Strategies for the detection of single nucleotide polymorphisms using the polymerase chain reaction

Waterfall, Christy Marie

Award date:
2002

Awarding institution:
University of Bath

Strategies for the detection of Single Nucleotide Polymorphisms using the Polymerase Chain Reaction

Submitted by
Christy Marie Waterfall
for the degree of PhD
of the University of Bath
2002

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis is dedicated to Berry.

Acknowledgements

I would like to thank Dr John Old (National Haemoglobinopathies Reference Service, Oxford, UK) and Dr Steve Keeney (Molecular Medicine Laboratory, Manchester Royal Infirmary, UK) for supplying the DNA samples that enabled effective validation of the SNP genotyping assays developed during this research. Thanks also to Action Research UK and Promega UK for kind permission to reproduce Figures 1.2 and 2.1.

My supervisors, Dr Benjamin Cobb and Dr Mark Enright deserve huge thanks for their help and encouragement throughout. In addition, thanks to Dr Michael Bennett for taking the time to discuss and proofread my work, Prof. Robert Eisenthal for expert advice regarding enzyme kinetics, and everyone at Molecular Sensing plc for contributing to a really enjoyable three years.

Mum and Dad deserve extra special thanks for their continued encouragement and support throughout my education- this is hugely appreciated and will never be forgotten. Last but not least, extra special thanks to Ben who has been wonderfully supportive throughout, giving me the confidence to believe that anything is possible.

Title i

Acknowledgements iii

Contents iv

Abbreviations vii

Abstract x

Chapter One

General Introduction

1.0 Genetic Variability 1

1.1 Targets for SNP Screening 7

1.2 Applied Genomics 11

1.3 Detecting Genetic Variability 14

1.4 Real-Time Quantitative PCR Methods 55

Chapter Two

Materials and Methods

2.0 PCR amplifications 65

2.0.1 General protocol using an MJ PTC-100 65

2.0.2 General protocol using a Roche LightCycler™ 65

2.0.3 Melt curve analysis using a Roche LightCycler™ 66

2.1 Plasmid DNA Library Construction 67

2.1.1 Preparation of plasmid constructs with a SNP 67

2.1.2 Ligation of PCR fragments into plasmid vector 67

2.2 Preparation and Transformation of Fresh or Frozen Competent *Escherichia Coli* cells 70

2.2.1 Preparation of fresh or frozen competent E. coli cells 70

2.2.2 Transformation of fresh or frozen competent E. coli cells 71

2.2.3 Testing bacteria for Insertional inactivation 72

2.3 DNA Purification 73

2.3.1 Wizard® PCR preps for DNA purification 73

2.3.2 QIAprep Spin Miniprep Kit Protocol 74

2.4 Plasmid DNA Quantitation 75

2.4.1 Agarose gel electrophoresis 75

2.4.2 Spectrophotometric analysis 75

2.4.3 Quantitative PCR (qPCR) methods 76

2.5 Human genomic DNA Isolation and Restriction 77

2.5.1 Isolation from buccal cell swab 77

2.5.2 Requesting DNA samples from external laboratories 78

2.5.3 DNA restriction, agarose gel electrophoresis 78

Chapter Three

Optimisation of Allele-Specific PCR (AS-PCR) for SNP Genotyping

3.0 Introduction 80

3.1 Results 94

3.1.1. Conventional allele-specific PCR tests 94

3.1.2 Single tube genotyping of sickle cell anaemia using bi-directional PCR 94

3.1.3 Reaction components affecting product yield 107

3.2 Discussion 110

Chapter Four

Rapid SNP genotyping using Fluorescent Bi-directional PCR and Melt Curve Analysis

4.0 Introduction 115

4.1 Results 119

4.1.1 Assay design 119

4.1.2 PCR amplifications 125

4.1.3 Assay optimisation 125

4.1.4 SNP genotyping 145

4.1.5 Genotype validation 152

4.2 Discussion 154

Chapter Five

Bi-directional PCR Combined with 5' Nuclease Fluorogenic Probes for Allelic Discrimination

5.0 Introduction 159

5.1 Results 166

5.1.1 Assay design 166

5.1.2 Assay optimisation and PCR amplification 170

5.1.3 SNP genotyping 176

5.2 Discussion 180

Chapter Six

Kinetic Characterisation of PCR: Application to Primer Mismatches and Quantitative PCR

6.0 Introduction 185

6.1 Results 190

6.1.1 Allele-specific PCR amplifications 190

6.1.2 Collection of 'per cycle' rate data 190

6.1.3 Quantitative PCR amplifications 198

6.2 Discussion 213

6.2.1 Development of kinetic model of PCR 213

6.2.2 Determination of kinetic constant in allele-specific PCR 216

6.2.3 High substrate inhibition 219

6.2.4 Absolute PCR quantification using external standards 220

Chapter Seven

General Discussion 224

References 229

Appendix I 255

Primer sequences for PCR amplifications

Appendix II 257

pGEM-T® Easy Vector System II plasmid sequence with cloned insert

Appendix III 261

Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis

Appendix IV 270

Rapid SNP genotyping using fluorescent bi-directional PCR

Appendix V 276

Locked Nucleic Acids as enhancers of allele-specific PCR

Appendix VI 279

Kinetic characterisation of mismatches in allele-specific PCR: A quantitative assessment

Abbreviations

A	adenine
ADR	adverse drug reaction
AGE	agarose gel electrophoresis
APEX	arrayed primer extension
ARMS	amplification refractory mutation system
ASA	allele-specific amplification
ASO	allele specific oligonucleotide
AS-PCR	allele-specific polymerase chain reaction
ATP	adenosine triphosphate
bp	base pair
C	cytosine
CCR	combined chain reaction
C_P	crossing point
C_T	cycle threshold
ddNTP	dideoxyribonucleotide triphosphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double stranded deoxyribonucleic acid DOL
EDTA	ethylenediaminetetraacetic acid
FAM	6-carboxyfluorescein
FP	fluorescence polarisation
FRET	fluorescence resonance energy transfer
g	gram
G	guanine
HbA	normal haemoglobin
HH	hereditary haemochromatosis
HbS	sickled haemoglobin
HTP	high throughput
JOE	6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein

kb	kilobase, 1000 base pairs
LB	Luria-Bertani
LCR	ligase chain reaction
LD	Linkage disequilibrium
LNA	locked nucleic acid
MALDI-TOF	matrix-assisted laser desorption-ionisation time-of-flight
MES	2-[N-morpholino]ethanesulphonic acid
mRNA	messenger RNA
ng	nanograms
OD	optical density
PASA	preferential amplification of specific alleles
PCR	Polymerase Chain Reaction
PNA	peptide nucleic acid
PPi	inorganic pyrophosphate
QC-PCR	quantitative competitive polymerase chain reaction
qPCR	quantitative polymerase chain reaction
RCA	rolling circle amplification
RFLP	Restriction fragment length polymorphism
Rn	reaction
RNA	ribonucleic acid
ROX	6-carboxy-X-rhodamine
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecylsulphate
SNP	Single nucleotide polymorphism
ssDNA	single stranded deoxyribonucleic acid
STR	Short tandem repeat
TAMRA	6-carboxy-N,N,N',N'-tetramethylrhodamine
TFB	transformation buffer
T	thymine
TET	tetramethylrhodamine
T_m	melting temperature
Taq	*Thermus acquaticus*
UV	ultra violet

Abstract

PCR is widely used as a research tool for SNP detection because it offers a simple, sensitive and robust platform. Its wider use within clinical diagnostics and drug development requires the development of assay formats that are inexpensive, rapid, and ideally use a single-tube for the detection of a bi-allelic SNP. A single tube adaptation of allele-specific PCR that identifies allele-specific amplicons according to size was developed for detection of the SNP responsible for the sickle cell anaemia phenotype. Combined with an effective optimisation strategy, previously described for maximising yield in PCR tests, this format reduces the reagent costs associated with running the test whilst improving throughput.

The assay was further rationalised to provide a homogenous reaction format that had significant cost and throughput benefit. Incorporation of the universal double strand DNA specific binding dye, SYBR® Green I, into the PCR reaction mixture enabled real time detection of amplification and identification of allele-specific products using melt curve analysis. A further adaptation of this assay involved the use of sequence-specific real time PCR probes to detect specific PCR products, introducing semi-sequence specificity to the test. The utility of the test was demonstrated using different real-time PCR instruments with a range of detection modalities and reagent systems.

Allele-specific PCR methods are currently the method of choice for routine laboratory analysis of DNA samples for mutations linked to sickle cell anaemia, and hereditary haemochromatosis. These were the model systems used for the development of the assay formats described here. In a preliminary clinical trial, genotyping results for the detection of sickle cell anaemia were 100% concordant with RFLP data. For the detection of the two mutations associated with hereditary haemochromatosis, C282Y and H63D, results were 97% and 92% concordant with data from a reference laboratory. The mistyped samples were attributable to sample mix-ups during transitions between laboratories.

The detailed kinetic aspects of allele-specific PCR tests were investigated quantitatively using a novel model describing the PCR process. This model assumes that PCR becomes limited by high substrate, high substrate inhibition, and good degree of correlation was achieved. Kinetic parameters including V_{max}, K_M and K_i were estimated for extension of matched and mismatched primers, generating data directly applicable to the design of rapid-cycle PCR-based SNP detection tests. The model was also used to determine the cycle threshold in a PCR reaction curve, providing quantitative estimations of known and unknown template concentration. The model was validated by accurately estimating the concentration of twelve samples of unknown initial concentration by comparison to an external standard curve. This method exhibited equal accuracy to a widely used commercially available method.

Chapter One

General Introduction

1.0 Genetic Variability

The understanding of genetic information has witnessed a huge progression over the past 100 years, starting with the first established cellular basis of heredity: the chromosomes and genes (Morgan, 1910). In 1953, Watson and Crick deciphered the molecular basis of this heredity by determining the double helical structure of DNA (Watson and Crick, 1953). This enabled the elucidation of biological mechanisms by which the cells read information contained in genes and the development of molecular biology methodologies that have allowed scientists to repeat these systems *in vitro*. Now, we are in the era of sequencing genes, and entire genomes to enable further advancements in the understanding of molecular mechanisms. To date, the genome sequences of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, seven archae, one fungus, one plant, and two animals, including the human genome have been elucidated (The International Human Genome Sequencing Consortium, 2001).

One of the greatest challenges for molecular biology post the Human Genome Project is to map the extensive genetic diversity that exists between individuals, and between species. The most abundant source of this variation is the randomly distributed single base changes in DNA sequence known as single nucleotide polymorphisms (SNPs). Recent studies have identified as many as 1.42 million single nucleotide polymorphisms distributed throughout the genome at an average density of 1 SNP every 1.9 kilobases (The International SNP Map Working Group, 2001). This dense set of highly stable markers opens up the possibility of association studies, identifying markers important in generating complex genetic traits. For example, disease loci may be pinpointed along with 'functional variants' in which the nucleotide change alters the function or expression of a gene that directly influences a disease state or drug response (Ingram, 1957). The study of the distribution of SNPs, particularly between different populations, is also a valuable

tool for investigating molecular events that underlie evolution, namely genetic drift, mutation, recombination and selection (Freeland and Hurst, 1998).

SNPs are defined as single base pair positions wherein different sequence alternatives (alleles) exist between individuals within a single population, or between individuals within different populations. Such polymorphisms must have an allelic frequency of 1% or greater within the test population to be considered as SNPs. Polymorphisms with lower frequencies are generally classified as 'rare mutations' (Brookes, 1999), although the definition used in this thesis encompasses disease predisposing single base variants of low frequency, or risk associated alleles as well.

Depending on the locality of a SNP within certain sequences, it can have different effects at the phenotype level. Generally, phenotype-influencing SNPs are termed 'functional' (nonsynonymous or missense) SNPs since they result in altered amino acids in the primary protein sequence. SNPs that do not alter the resulting amino acid due to the degeneracy of the genetic code are termed synonymous SNPs, e.g. a change of CCA mRNA codon to CCG does not change the designation of the amino acid proline (Lewin, 1995).

SNPs occurring in the coding regions of genes (cSNPs) have been identified as the root cause of a number of recessively and dominantly inherited disorders (Section 1.1), and different drug responses in individuals by altering the primary structure of proteins involved in drug metabolism (Section 1.2; Ginsburg and McCarthy, 2001). More commonly, SNPs interact with environmental factors and other genetic variants from multiple genes causing complex disorders such as late onset Alzheimer's disease (linked to mutations in the apolipoprotein E gene on chromosome 19, Davignon *et al.,* 1988; Wilson *et al.,* 1996), and thromboembolism (linked to the factor V Leiden mutation in the coagulation Factor V gene and the $PT_{G20210A}$ mutation in the prothrombin gene, Bertina *et al.,* 1994; Voorberg *et al.,* 1994). SNP markers are used today as the basis of clinical assays to screen individuals for key inherited diseases.

SNPs can also occur within sequences with regulatory function (rSNPs) and within splice sites. Potentially these SNPs can disrupt gene expression and so contribute to an increased risk of developing certain diseases. For example, SNPs in the promoter regions of cytokines, namely tumour necrosis factor (TNF), interleukin 4, interleukin 6 and interleukin 10 (e.g. genes that coordinate the immune response), have been associated with a range of infectious and autoimmune disorders, including tuberculosis (Cooke and Hill, 2001) and arthritis (Taylor *et al.,* 2001). This trend suggests that small differences in the regulation of key genes by the presence of a SNP can amplify or suppress a biological pathway such as a host response to a pathogen, causing the diversity in response to such disorders (Modiano, 1996).

Approximately 60,000 functional SNPs have been reported (The International SNP MAP Working Group, 2001; Pirmohamed and Park, 2001), these are a small fraction of the 1.42 million total SNPs identified. The remaining SNPs occur in non-coding regions of the genome and are not associated with known phenotypes (Taylor *et al.,* 2001).

SNPs have long been the marker of choice for studies in genetic variation. Initially, these polymorphisms were known as 'restriction fragment length polymorphisms' (RFLPs). In this form, SNPs were identified according to their ability to create or abolish scission sites for restriction enzymes (Murayama, 1966). Subsequently, SNPs as short tandem repeats (STRs) or microsatellites replaced RFLPs as markers for variability studies (Botstein *et al.,* 1980) due to their high level of allelic variation but, since the completion of the human genome sequence, SNPs have been rediscovered as the marker of choice for variability studies. They provide a highly abundant genetic marker with greater stability than STRs due to their low mutation frequency. SNPs have an additional advantage in that STR loci are markers in the sense that the polymorphism in the STR is used to locate an adjacent functional variant that contributes to the disease state; variation in the STR itself rarely has a direct causative effect on the phenotype. Like STRs, SNPs can be used as a marker for a nearby functional variant, but since SNPs can occur in the coding or regulatory regions of genes, they can have direct functional consequences (Gray *et al.,* 2000).

SNPs can also be inherited in highly conserved, interactive blocks called 'haplotypes'. A particular pattern or combination of sequential SNPs found on a single chromosome that tend to be inherited together with little genotypic variation is often referred to as a haplotype (haploid genotype). The study of SNP haplotypes may help streamline the analysis of genetic factors responsible for complex diseases, because without prior knowledge of the functional variant, genomic regions can be evaluated for association with disease outcomes. Subsequently, the functionally significant SNPs can be identified by a closer examination of regions identified by haplotype-based studies (Taylor *et al.,* 2001; Emahazion *et al.,* 2001).

The two principal strategies used currently by researchers for genetic association studies are 'linkage analysis' and 'association analysis' (or 'candidate gene approach'), summarised in Figure 1.1 (Taylor *et al.,* 2001). Linkage analysis relies on linkage disequilibrium (LD) or non-random pattern of co-inheritance between genetic markers within close proximity to each other. Generally, familial-based studies are employed in which disease inheritance can be compared to patterns of linkage over large genomic regions. Through LD, associations found with anonymous markers can identify a region of the chromosome that may harbour a susceptibility gene without any *a priori* assumptions about what or where the susceptibility gene is (Kruglyak, 1999). However, linkage analysis rarely uncovers the specific mutation responsible for the disease outcome, rather a genetic marker within megabase resolution of where the disease gene is located. In addition, for human studies it is often difficult to collect multi-generation families of sufficient size or number to successfully apply these methods (McCarthy and Hilfiker, 2000).

For effective linkage, the mutation that is sought after must be a strong predictor of the disease phenotype, ideally necessary and sufficient to cause the disease. Linkage methods have enabled the identification of monogenic disorders because rare disease alleles show strong linkage with surrounding markers (Nowotny *et al.,* 2001). However, linkage analysis is not optimal in the detection of genes responsible for the aetiology of complex disorders such as Alzheimer's disease, thromboembolism, osteoporosis (Giguére and Rousseau, 2000), cardiovascular

and inflammatory diseases (Humphries *et al.,* 2001), psychiatric diseases and most cancers (Kim *et al.,* 2000; Bennett *et al.,* 1999) where it is often the combined effect of several SNPs in key genes and environmental factors that determine whether an individual develops a disease (Allen *et al.,* 1988).

Association studies are the preferred route for mapping complex disease genes because they use *a priori* knowledge of biological function to nominate candidate genes that are likely to have a role in the disease aetiology (Emahazion *et al.,* 2001). Typically, association studies use a case-control or cohort strategy involving unrelated subjects, overcoming the problems encountered in gathering enough subjects for family-based linkage studies. If a SNP or genetic variation is more prevalent in patients with a particular disease compared to unaffected controls, an association is inferred. In principle, sequences conferring only a fractional increase in disease risk can be detected, and the actual marker involved in the disease can be located, as well as genetic markers that are themselves in linkage disequilibrium with alleles of the associated marker (Taylor *et al.,* 2001).

The number of SNP markers required to detect 'functional' single gene variants within large populations is greater for effective linkage studies than candidate gene analysis (Kruglyak, 1999). Estimates vary from 1 million to 30,000 SNPs in at least 1000 individuals, which realistically exceeds the capabilities of current technologies (Syvänen, 2001). The use of pooled samples, comprised of equal amounts of genomic DNA from up to 1000 individuals, is one of the strategies employed as a means of reducing the number of genotyping reactions required (Breen *et al.,* 2000; Germer *et al.,* 2000; Chen *et al.,* 2002). The candidate gene approach represents a more achievable target given that smaller numbers of SNPs are required and the scale is no longer random and genome-wide. For example, the genotyping of 20 SNPs in 50 candidate genes in 1000 samples represents a more reasonable goal of 1 million genotypes. Both strategies would benefit from improved genotyping technologies with lower overall costs and increased throughput.

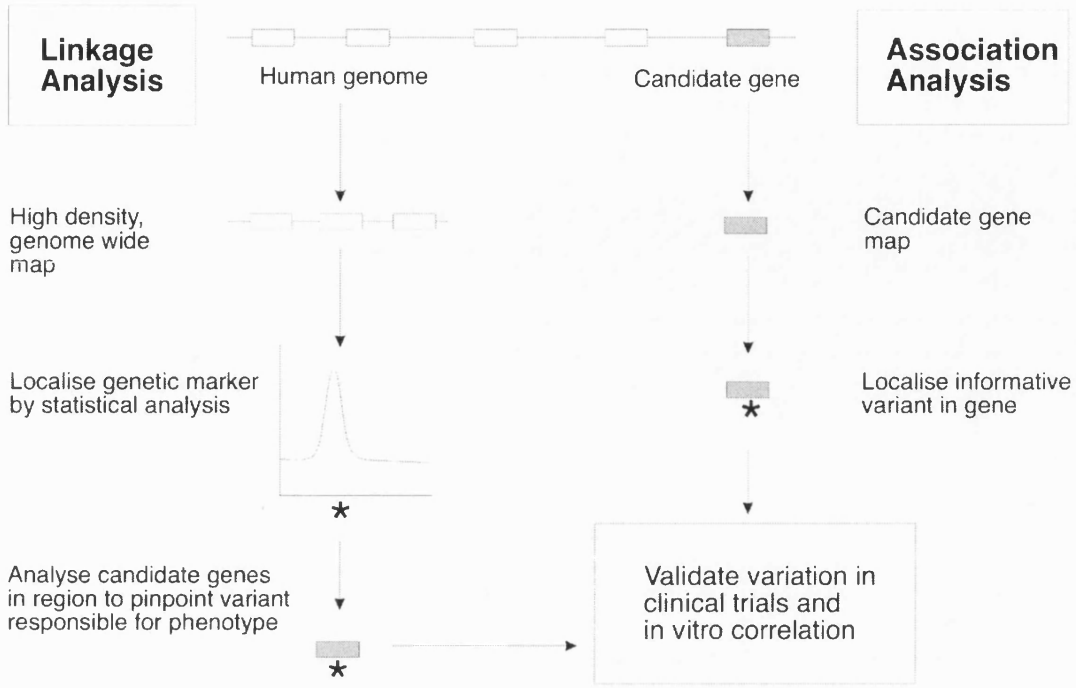


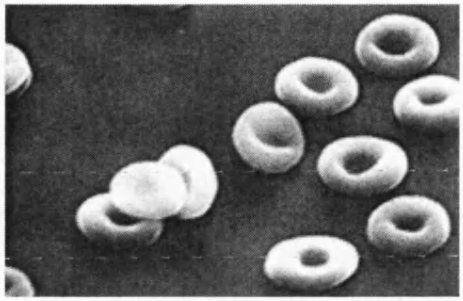
Figure 1.1: Two principle approaches to associating genetic variation with contribution to disease cause and susceptibility. In the Linkage analysis method, a genome-wide association approach is used, which requires a dense set of genetic markers (SNPs or STRs) distributed across the genome. A genetic marker is localised to a region of the chromosome if a sufficiently high correlation is observed. That region is then scanned to locate the actual susceptibility variation. In the Association analysis (or candidate gene approach), candidates SNPs are selected on the basis of biological knowledge or genetic observations. The candidate gene only is mapped, reducing the scale of the initial scan. *Denotes informative genetic marker.

1.1 Targets for SNP Screening

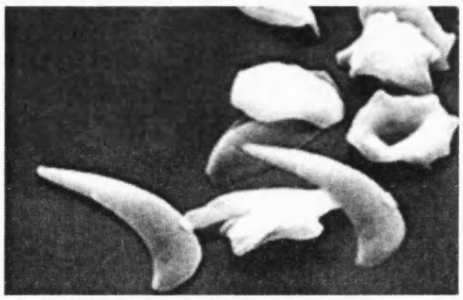
Approximately 1600 diseases characterised in the Online Mendellian Inheritance in Man (OMIM) database are caused by a single point mutation (monogenic disorders). These represent targets for genomic screening in addition to the thousands of characterised SNPs underlying common multifactorial disorders and variabilities in drug response. In this work, three clinically significant polymorphisms discussed below are used as 'model SNPs' for rapid genomic screening.

Sickle cell anaemia (or β-thalassaemia major) was the first trait to reveal that a single point mutation can change protein structure and lead to a disease phenotype (Ingram, 1957). At the molecular level, the sickle cell anaemia mutation is characterised by a single A→T missense transversion in the sequence encoding codon 6 of the human β-globin gene causing the substitution of the amino acid glutamine to valine (Murayama, 1966; Bookchin *et al.,* 1970). This single molecular defect forms a mutant haemoglobin chain (known as HbS), which in the deoxygenated state has a tendency to polymerise and aggregate (Figure 1.2). The polymerisation is thought to cause a reversible increase in red cell (erythrocyte) membrane permeability that then triggers a chain of events where two constitutive transporters of the erythrocyte membrane become activated (a Ca^{2+}- sensitive K channel and an electroneutral K:Cl cotransporter) leading to sickle cell dehydration. The extent of red blood cell dehydration and density has been correlated with the clinical severity of the disease (Bookchin and Lew, 2002).

Physiologically, the sickled erythrocytes become trapped in the microcirculation, depriving organs of essential oxygen, causing pain and chronic anaemia. The life span of sickled erythrocytes in the circulation is significantly shortened to 10-12 days in most patients compared to normal erythrocytes (120 days). Sickle cell carriers (heterozygotes) are clinically normal, although their erythrocytes will sickle when subjected to very low oxygen pressure *in vitro* (Bookchin and Lew, 2002).



Normal red blood cell



Sickled red blood cell

Figure 1.2: Normal and sickle cells. Electronmicrographs reproduced with kind permission from Dan Sartin at Action Research UK (www.actionresearch.co.uk).

Homozygous HbS is common among, but not confined to, peoples of African American origin. Individuals in parts of Europe, South America and the Middle East are also commonly affected. In the UK, sickle cell disease affects approximately 10,000 people (Bookchin and Lew, 2002), some 3,000 more individuals than cystic fibrosis, which has regular screening programmes in place (Davies *et al.,* 2000). In certain areas of the UK, individuals are tested for sickle cell anaemia depending upon their ethnicity. With increased immigration and mixed race parenting, ethnicity may no longer be a sufficient criterion on which to select screening candidates. The aim is to implement a more widespread screening program for this disease in the UK (Davies *et al.,* 2000), with a view to preventing morbidity and mortality by the administration of penicillin, in addition to genetic counselling programmes for the parents of affected children (Serjeant, 1997).

Hereditary haemochromatosis (HH) is the second disease selected as a target for genomic screening in this thesis. HH is one of the most common human genetic diseases, characterised by iron overload, in which iron is mostly deposited in parenchymal cells in the liver, pancreas, heart, and other organs causing malfunction (Andrews, 1999). At the molecular level, two single point mutations are associated with the development of HH. The first and more clinically significant (occurring in 90% of HH patients; Townsend and Drakesmith, 2002) is a G-to-A transition at nucleotide position 845 of the open reading frame of the haemochromatosis (*HFE*) gene. The resulting HFE protein exhibits a change from cysteine to tyrosine at amino acid position 282 (known herein as C282Y). The homozygous C282Y mutation found is 90% of patients with HH. Heterozygous inheritance of C282Y also exhibits minor abnormalities in iron overload (Townsend and Drakesmith, 2002).

The link between the C282Y mutation and the iron overload characteristics of HH has been hypothesised, based on findings that the C282Y mutation inhibits HFE/β2-microglobulin binding. In normal iron metabolism, association of HFE to β2-microglobulin is necessary for intracellular trafficking and incorporation of the HFE protein in cell membrane. The HFE protein then associates with the transferrin

receptor at the cell surface, inducing activation of the transferrin receptor by phosphorylation. This activation reduces receptor affinity and impairs endocytosis of iron-saturated transferrin, with a decrease in cellular iron uptake as a result (Carlson and Olsson, 2001). The C282Y mutation is proposed to remove an essential cysteine residue, which normally participates in a disulphide bond, forming a structural conformation of the HFE protein that is capable of interaction with β2-microglobulin. The disruption of this important cell signalling interaction means that transferrin receptor affinity for iron-saturated transferrin is not reduced when required, and iron overload ensues (Feder *et al.,* 1997).

The second mutation in the *HFE* gene, H63D, can cause the disease when inherited with a single C282Y mutation (the so-called 'compound heterozygote' state). H63D represents a C-to-G transversion at nucleotide position 187 of the open reading frame of the *HFE* gene, causing a change from histidine to aspartic acid at amino acid position 63 in the resulting HFE protein (Ugozzoli *et al.,* 2002). Increased susceptibility to HH is observed in 38% of compound heterozygotes. The specific role of the H63D mutation may be in altering the affinity of the HFE protein for another ligand involved in iron metabolism, or alter the manner in which the mutant protein interacts with other proteins in the cell membrane (Feder *et al.,* 1997), although the exact role is, as yet, unproven.

The C282Y and H63D mutations are common in people of Northern European origin, affecting approximately 1 in 200-300 people, and more prevalent than sickle cell anaemia among the African American population (Baty *et al.,* 1998). Morbidity and mortality in individuals affected with HH could be significantly reduced if the disease were diagnosed prior to the onset of hepatic, cardiac, or endocrine dysfunction. Therefore, screening of these SNP markers is important because an effective therapy is available through periodic phlebotomy to remove excess iron in the blood (Carlsson and Olsson, 2001).

1.2 Applied genomics

It is widely anticipated that the study of SNP markers will also provide the ability to understand the *"cause and effect"* of individual variation, allowing prediction of disease susceptibility or drug response. SNP genotyping is already performed on a routine basis to screen for monogenic diseases, as well as "susceptibility" genes in the complex diseases. Increasing interest surrounds the use of SNP markers to predict an individual's response to therapeutic treatment, which may facilitate preventative medicine or the tailoring of specific treatment according to an individual's predisposition to disease (Wieczorek and Tsongalis, 2001).

This informed approach to "personalised medicine" is broken down into two main disciplines: *pharmacogenomics*, the hereditary differences that dictate drug behaviour, and *pharmacogenetics*, the variability in drug response and metabolism due to hereditary differences (Ginsburg and McCarthy, 2001). Basically, pharmacogenomic studies involve the analysis of a single drug, and its effect on a population of genomes (people), whereas pharmacogenetics involves the study of a single genome (person) and the effects of many drugs on the individual (Klaus Linderpainter, Personal Communication 2002). Confusingly, the terms are generally used interchangeably.

The two parallel approaches described for disease analysis (Figure 1.1, pg. 6), can also be applied to map the variety of individual responses to therapeutics. Although age, diet and physiological factors are important, much of the variation observed in drug response between patients is thought to be attributable to the underlying genetic variation of the human population (Wieczorek and Tsongalis, 2001). It is hoped that pre-treatment screening for these alleles will allow doctors to predict a patient's response to therapy, improving the safety and efficacy of therapeutics by avoiding adverse drug reactions (ADRs). ADRs remain a major clinical problem, with some estimates putting ADR as the fourth cause of death in the US, representing US$75 billion in healthcare costs annually. Also, ADRs are considered

one of the most common factors that contribute to drug withdrawal, representing a significant financial burden to the healthcare industry (Pangas and Woodruff, 2002).

As a result, a large amount of research has been undertaken to investigate links between SNPs and drug response, revealing that the majority of polymorphisms identified so far are located in the drug metabolising enzymes, or drug receptors and transport proteins (Liljedahl and Syvänen, 2002). Today routine pharmacogenetic analyses focus on typing drug metabolising genes belonging to the cytochrome P450 (CYP) gene family, which contain many of these polymorphisms. Over 30 different isoforms of cytochrome P450 enzymes exist, and it is estimated that more than half of all currently used drugs are metabolised by these enzymes. Polymorphisms in P450 enzymes have been shown to lead to poor, normal or ultra-rapid metabolism of drugs. For example, SNPs in the CYP2D6 gene affect individual responses to a large number of commonly used drugs such as β-blockers, antidepressants, and codeine. Also, a poor metabolising variant of the enzyme encoded by the CYP296 gene causes life-threatening bleeding in individuals exposed to the anticoagulant warfarin (Furuya *et al.,* 1995).

To improve efficacy and general acceptance, drug developers try to avoid compounds whose metabolic pathways are significantly influenced by polymorphisms in the P450 enzymes (Pirmohamed and Park, 2001). By moving on from the current pharmaceutical model of "one-size-fits-all" prescription of drugs, and moving towards a tailored prescribing of "personalised medicine" using pharmacogenomic techniques, more drugs may be moved forward in the drug discovery pipeline.

Elucidation of these pharmacogenetic effects will have beneficial consequences in many other areas of healthcare (Figure 1.3). With the availability of clinical tests for key SNPs, the options for drug developers will increase, not only determining the appropriate drug to be prescribed, but also the optimal dosage to administer. To satisfy the potential volume of genotyping tests that such a system would require, improved assays with reduced costs and improved throughput are a necessity.

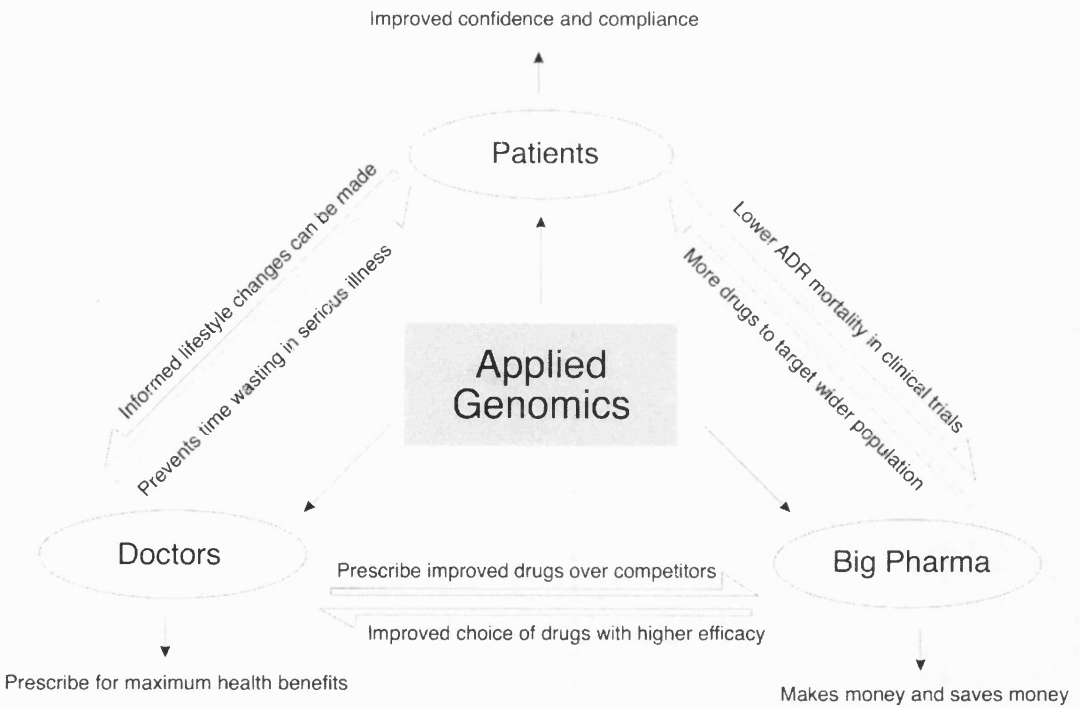


Figure 1.3: A schematic outlining the healthcare areas benefiting from applied genomics (pharmacogenomic and pharmacogenetic information). Information extracted from (Wieczorek and Tsongalis, 2001; Pirmohamed and Park, 2001).

1.3 Detecting Genetic Variability

SNP genotyping (SNP analysis) has become a generalised term encompassing a broad range of methods used to detect genetic variability within individuals. The three main stages of SNP analysis are SNP discovery (or SNP identification), SNP validation, and SNP scoring (genotyping, calling, or typing; The International SNP Map Working Group, 2001).

For SNP discovery, the aim is to screen SNPs in DNA samples from as many individuals from as many populations as possible, ideally with an ultra-high throughput solution. Currently, Sanger's dideoxy method for DNA sequencing (Sanger *et al.,* 1977), as used in the Human Genome Project (Lander *et al.,* 2001; Venter *et al.,* 2001), provides a large screening capacity. To improve throughput parallel sequence reactions' on multiple sequence platforms were used. Although sequencing remains the gold standard method of SNP discovery (The International Human Genome Sequencing Consortium, 2001), it is relatively slow and expensive on a large scale.

To validate a SNP, confirming that a candidate SNP is not the product of experimental error, a gene harbouring a 'candidate SNP' is generally re-sequenced. A SNP map location (actual position within the human genome) and allele frequencies are then assigned. To increase throughput for SNP validation, high-density microarray technology is often used allowing the genotyping reactions to be multiplexed. Typically, a microarray carries tens or even hundreds of oligonucleotide probes for each SNP to be analysed immobilised on a solid platform (e.g. GeneChip from Affymetrix). The probes include all possible sequences at the SNP site, and a stretch of nucleotide sequence that flanks the SNP. Specialised software is used to interpret the patterns arising from a matched hybridisation to assign the genotypes (Lipshutz *et al.,* 1995; Syvänen, 2002).

The final stage, SNP scoring, has been described as the "post-genomic" era of SNP analysis, since it utilises the genomic information revealed by discovery and

validation stages discussed above (Kwok, 2001). SNP scoring uses the known DNA sequence information as a basis on which to design assays for the detection of a characterised genetic variation within the genomic sequence. Usually the variant has been associated with a particular clinical phenotype. Effective detection of genetic variability within individuals requires assays that are inexpensive and rapid without compromising on accuracy, even in sub optimal conditions. In addition, the test should involve minimal manipulation to reduce the opportunity for sample mix-ups and contamination. The key attributes sought in the development of clinical genotyping assays for SNP scoring is summarised in Figure 1.4.

Underlying any genotyping assay is a biochemical principle that is used to locate a SNP generating a 'detectable' signal in its presence. The most important biochemical principles employed in SNP-genotyping platforms today are summarised in Figure 1.5. In general, they employ hybridisation with sequence specific oligonucleotide probes, or DNA modifying enzymes, such as DNA polymerases and ligases (Syvänen, 2001). These biochemical reactions have been multiplexed, automated and adapted to suit a range of detection methods to identify the specific biochemical signal generated in the presence of a target nucleic acid sequence. The most common detection modality is monitoring light emission since there are many ways to do so, including direct fluorescence (Figure 1.6A), fluorescence resonance energy transfer (FRET; Figure 1.6C), fluorescence polarisation (FP; Figure 1.6D) and luminescence (Figure 1.6 E). Mass spectrometry is becoming increasing popular since it offers the most direct method of detection i.e. molecular weight measurement. It also takes just milliseconds to analyse each sample, and moderate multiplexing is possible. However, the analyte needs to be of very high purity to eliminate a noisy signal, which has been the main drawback to date (Figure 1.6 B; Jackson *et al.,* 2000). Novel methods exploiting the electrochemical properties of DNA for SNP detection have also been described in which oligonucleotide probes are deposited on an electrode. The electrical property of the probe is altered when the complementary DNA is annealed to it (Figure 1.6 F; Kelley *et al.,* 1999). Some refining of electrochemical methods is required to achieve higher throughput before wider use can ensue (Kwok, 2001).

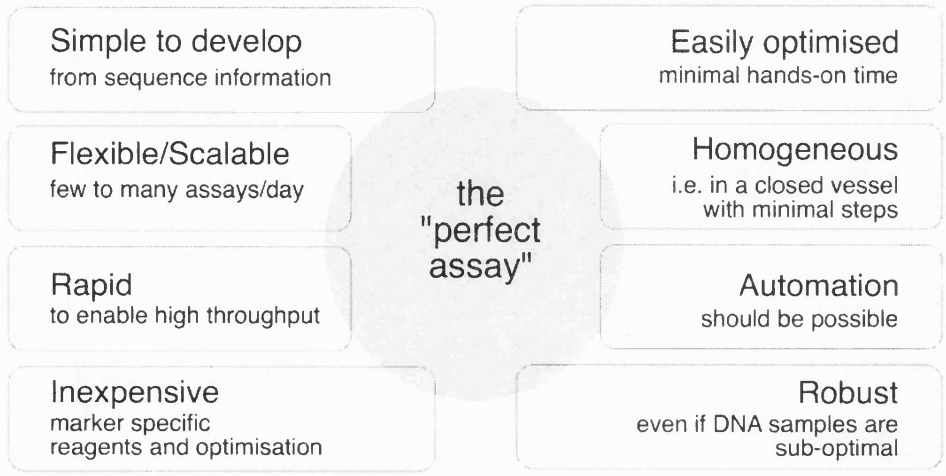


Figure 1.4: Key attributes to consider for the development of a "perfect assay" for SNP genotyping.

Figure 1.5: Six biochemical reaction principles underlying SNP genotyping assays (based on information from Kwok *et al.,* 2001, Syvänen, 2001). The grey panel demonstrates a C-to-A polymorphism. Probes shown are designed to detect the wild-type allele. [A] *Allele-specific hybridisation*: Usually, two oligonucleotide probes are used, one complementary to the wild type allele, and the other complementary to the mutant sequence. The variable position is usually designed to lie in the centre of the probe for maximal destabilisation in the presence of a mismatch. The probes will hybridise to target DNA under optimised conditions where a match is strongly favoured over a mismatch. [B] *Allele-specific primer extension:* Two primers are designed such that the 3' terminus ends on the mutation site. The 3' base of each primer is designed to match either the wild type or the mutant sequence. Under optimised conditions, *Taq* DNA polymerase will only extend a perfectly matched base. [C] *Single nucleotide primer extension (Minisequencing):* One primer that anneals to its target sequence immediately adjacent to the mutation site is extended by DNA polymerase with a single nucleotide that is complementary to the mutation site. The base at the mutation site can be inferred by identifying the nucleotide by which the primer was extended. [D] *Allele-specific oligonucleotide ligation:* Two adjacent oligonucleotides are included in a reaction with DNA ligase. The two oligonucleotides will be ligated together, only if the oligonucleotides perfectly match the template at the junction. The presence of a SNP can be inferred by determining whether ligation has occurred. [E] *Invasive cleavage (Invader assay):* Allele-specific oligonucleotide probes are designed base to match the wild type or mutant DNA sequence. The oligo sequence 5' of the mutation site is unrelated to the target. An upstream oligonucleotide (the Invader) is included, with complementarities to the 5' end of the allele-specific oligonucleotide. If an allele-specific oligonucleotide is perfectly matched to the target sequence, it is displaced by the Invader oligonucleotide, and the overlapping complex formed is specifically recognised by a FLAP endonuclease, which cleaves and releases the 5' end of the allele-specific oligonucleotide. Detection of the released 5' portion of the oligo will determine the nature of the allele present. [F] *Restriction digestion:* The presence of SNP may create or abolish a scission site for a particular restriction enzyme. Incubation of DNA with a restriction enzyme with known properties will either cleave or remain uncleaved. Determining the size of fragments following incubation will allow allele nature to be inferred.

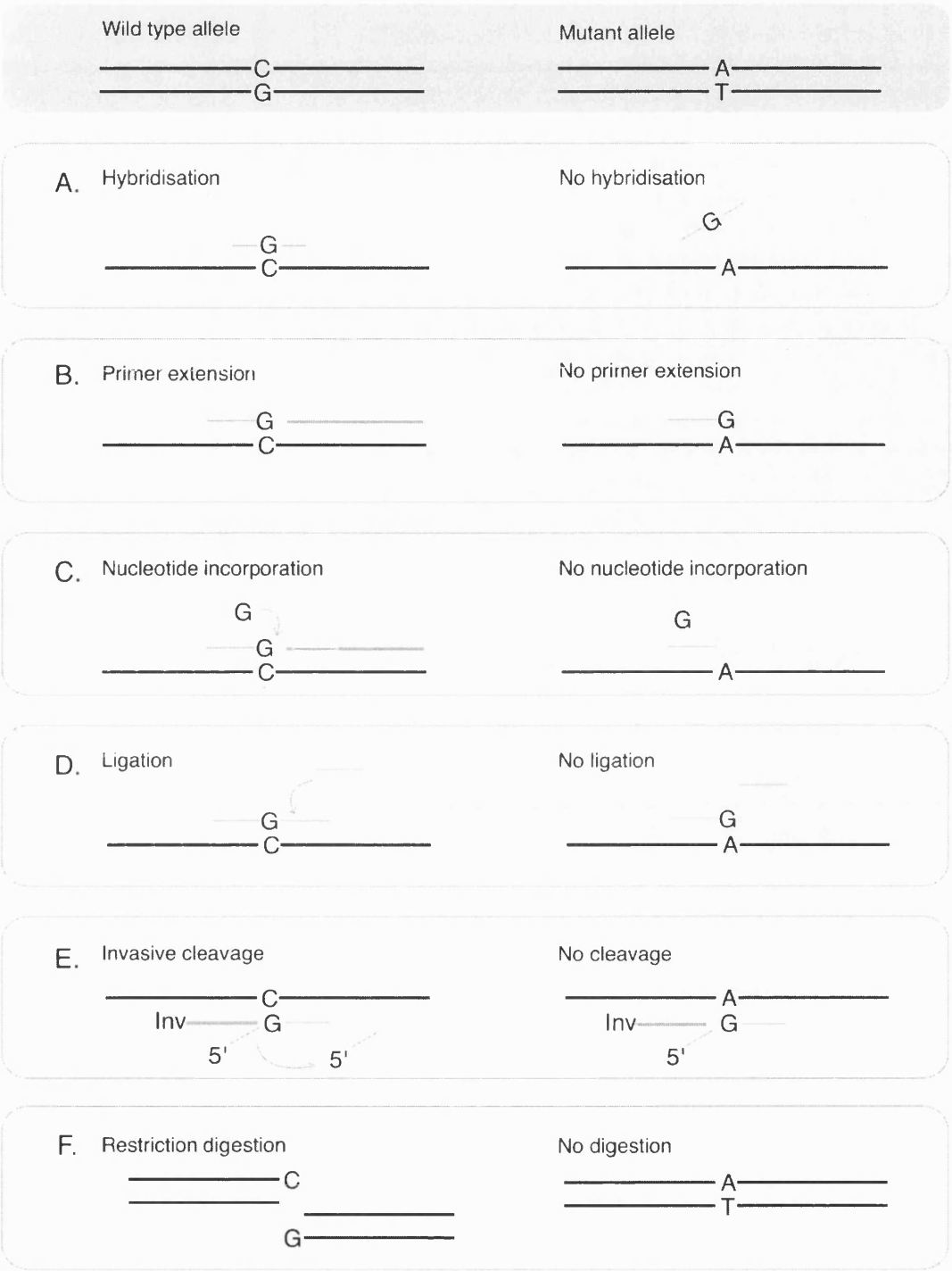
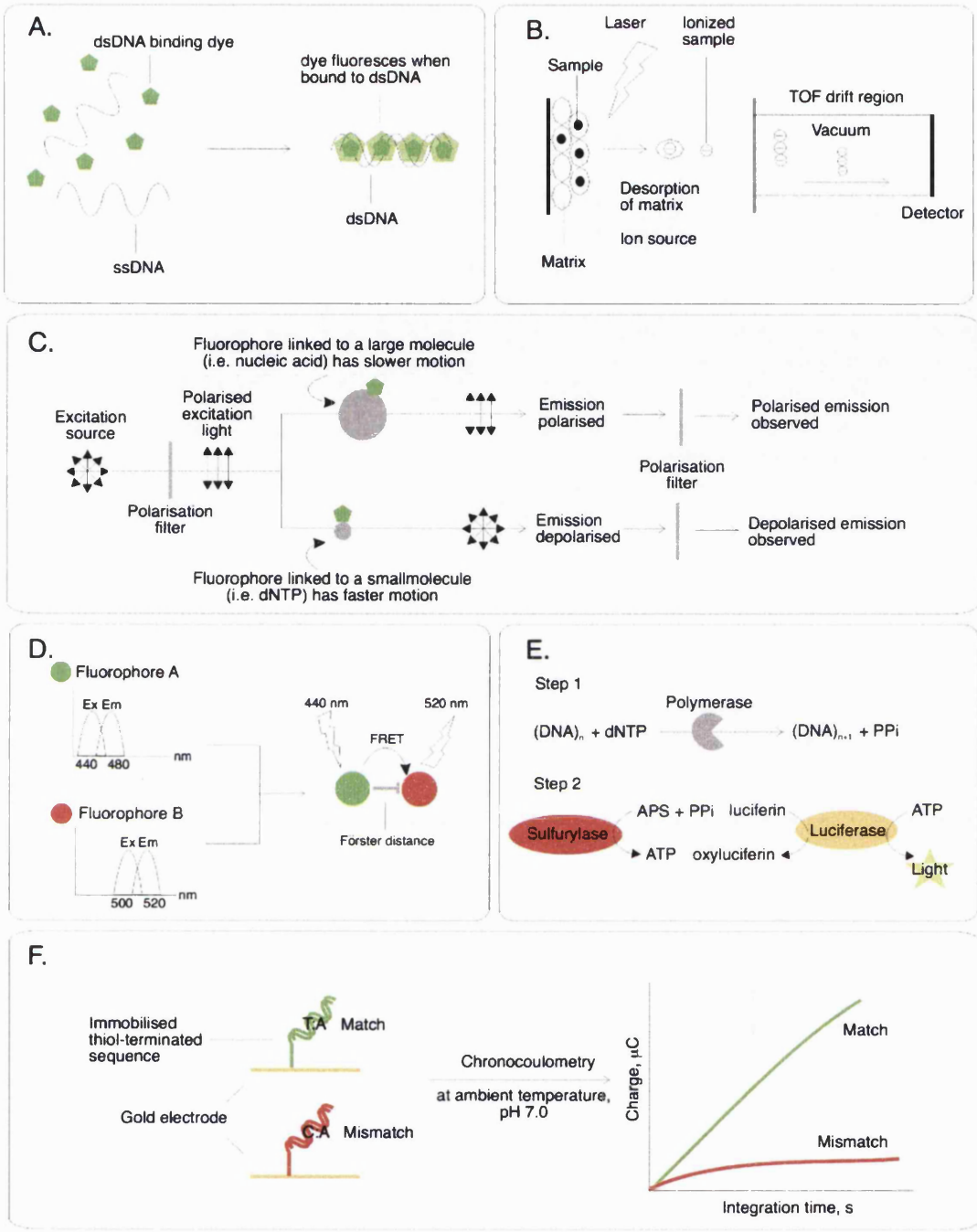

Wild type allele
C
G
Mutant allele
A
T
A.
Hybridisation
G
C
No hybridisation
G
A
B.
Primer extension
G
C
No primer extension
G
A
C.
Nucleotide incorporation
G
G
C
No nucleotide incorporation
G
A
D.
Ligation
G
C
No ligation
G
A
E.
Invasive cleavage
C
Inv
G
5'
5'
No cleavage
A
Inv
G
5'
F.
Restriction digestion
C
G
No digestion
A
T

Figure 1.6: The main detection modalities employed in genotyping assays today. The list is not intended to be comprehensive. [A] *Direct fluorescence* using dyes that specifically bind double-stranded DNA (dsDNA). Upon binding and excitation by a light source, an increase in detectable fluorescence is observed (Singer, *et al.,* 1997); [B] *Mass spectrometry*: example shows Matrix-assisted laser-desorption-ionization and time-of-flight (MALDI-TOF) where sample molecules are co-crystallised with the matrix and a laser beam is used to desorb and ionise the analyte. Desorption produces ionised sample molecules, which are accelerated by an electric field. The molecules travel along the TOF drift tube in a vacuum; the ions with larger mass-to-charge ratios travel more slowly than those with smaller mass-to-charge ratios. The time taken for ions to traverse the distance from source to the detector is accurately measured and equated to the mass-to-charge ratio of the ion (Jackson *et al.,* 2000); [C] *Fluorescence polarisation (FP)* is based on the observation that when a fluorescent molecule is excited by plane-polarised light, it emits a polarised fluorescent light into a fixed plane if the molecules remain stationary between excitation and emission (Perrin, 1926). At constant temperature and viscosity, FP is directly proportional to the molecular weight of the molecule to which the fluorophore is attached. If the fluorescent molecule is large (with high molecular weight), it rotates and tumbles more slowly in space and FP is preserved, providing a polarised signal at the detector. If the molecule is small (with low molecular weight), it rotates and tumbles faster and FP is largely lost, generating a depolarised signal at the detector (Chen *et al.,* 1999); [D] *Fluorescence Resonance Energy Transfer (FRET)* is a photophysical effect where energy that is absorbed by a donor fluorescent molecule (A) is transferred non-radiatively to an acceptor fluorescent molecule (B). Fluorophore A absorbs light at 440 nm and emits at 480 nm. Fluorophore B absorbs light at 500 nm and emits at 520 nm. When the fluorophores are brought within close enough proximity to each other (the Förster distance), energy transfer can occur. Fluorophore A Is excited by absorbing light of 440 nm and transfers the energy to fluorophore B. Now fluorophore A is excited and falls back to its ground state, thereby emitting light at 520 nm. Thus, when A and B are in close proximity the emission at 480 nm is decreased and the emission at 520 nm is increased (Stryer and Haugland, 1967); [E] *Luminescence*: Step 1 shows a polymerisation reaction catalysed by DNA polymerase, in which a dNTP is incorporated into the DNA strand, accompanied by the release of inorganic pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. Step 2 the quantitative conversion of PPi to ATP, by ATP sulfurylase in the presence of adenosine 5' phosphosulfate (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible lights in amounts that are proportional to the amount of ATP. Thus, each light signal is proportional to the number of nucleotides incorporated (Lundin *et al.,* 1976); [F] *Electrochemical detection:* A single example is shown here, DNA sequences containing a match or mismatch to the complementary probe were produced, and immobilise onto a gold electrode surface. Electrode surfaces modified with the duplexes were interrogated in chronocoulometric analyses at –350 mV with 0.5 μM methylene blue as the intercalated catalyst and 2.0 mM $[Fe(CN)_6]^{3-}$ as the solution-borne substrate. The diminished DNA-mediated electron-transfer efficiency as a result of local base stack perturbation caused by the mismatch is demonstrated by a decrease in charge signal (Boon *et al.,* 2000).


A.
dsDNA binding dye
dye fluoresces when bound to dsDNA
dsDNA
ssDNA
B.
Laser
Ionized sample
Sample
TOF drift region
Vacuum
Desorption of matrix
Detector
Ion source
Matrix
C.
Fluorophore linked to a large molecule (i.e. nucleic acid) has slower motion
Excitation source
Polarised excitation light
Polarisation filter
Emission polarised
Polarised emission observed
Polarisation filter
Emission depolarised
Depolarised emission observed
Fluorophore linked to a smallmolecule (i.e. dNTP) has faster motion
D.
Fluorophore A
Ex Em
440 480
nm
Fluorophore B
Ex Em
500 520
nm
440 nm
520 nm
FRET
Förster distance
E.
Step 1
Polymerase
(DNA)n + dNTP
(DNA)n+1 + PPi
Step 2
Sulfurylase
APS + PPi
luciferin
Luciferase
ATP
ATP
oxyluciferin
Light
F.
Immobilised thiol-terminated sequence
Gold electrode
T:A Match
C:A Mismatch
Chronocoulometry
at ambient temperature, pH 7.0
Charge, µC
Match
Mismatch
Integration time, s

The earliest methods of SNP scoring were developed for the diagnosis of sickle cell anaemia in the DNA of affected individuals. The genotyping method used was RFLP based on the linkage of the sickle cell allele to *Hpa* I restriction site (Figure 1.5F; Kan *et al.,* 1972). The single base alteration in the scission site of *Hpa* I gave a different DNA fragment pattern following digestion, than an unaltered site in unaffected individuals. Therefore, by visualising the resulting DNA fragment patterns, by electrophoresis, Southern blotting (Southern, 1975) and specific hybridisation to a cloned gene or oligonucleotide probes, the genotype was inferred. Later it was shown that the sickle cell mutation itself affected the cleavage site for *Dde* I (Geever *et al.,* 1981) and *Mst* II (Chang and Kan, 1982) and could be detected directly using enzymatic cleavage. Direct allele-specific oligonucleotide (ASO; Figure 1.5A) hybridisation methods were also described for the detection of the same mutation (omitting the RFLP step) in which differential hybridisation of allele-specific radioactive probes could be detected by autoradiography (Conner *et al.,* 1983). However, all of these methods are not widely used nowadays since they are laborious, difficult to implement, use radioactive probes (entailing complex safety precautions) and have very low sensitivity (Chehab *et al.,* 1987; Wu *et al.,* 1989).

It wasn't until the polymerase chain reaction (PCR) was invented in 1985 (Mullis and Faloona, 1986; Saiki *et al.,* 1985), that it became possible to design sensitive assays to detect SNPs in genomic DNA. PCR is a nucleic acid amplification technique in which short oligonucleotide sequences (primers), usually between 10 and 40 base pairs long, are designed to anneal to complementary sequences flanking either side of a 'target region' (Figure 1.7). These primers are added in excess to the target sequence DNA (or template). A suitable buffer, magnesium chloride ions, a thermostable polymerase and free nucleotides are also added. A process of thermal cycling is used to amplify the DNA several million-fold. The target DNA is initially denatured at 95°C and then cooled to generally between 40°C to 60°C to enable sequence-specific annealing of the primers to the separated strands. The temperature is then raised to the optimal temperature of the polymerase, generally 72°C, which then extends the primer, copying the target sequence. This series of events is repeated (usually 20 to 40 times), wherein during

the first few cycles, copies are made of the target sequence, and then during subsequent cycles, copies are made from copies, which increases target amplicon yield exponentially prior to DNA detection/visualisation (Figure 1.8).

Few technologies other than PCR have had such a revolutionary effect on how molecular studies are conducted (Erlich *et al.,* 1991). The impact on genotyping has been immense, since almost all genotyping performed today requires a PCR step of the region surrounding the mutation to provide the required sensitivity and specificity for distinguishing between heterozygous and homozygous genotypes.

In 1989, PCR itself was adapted for mutation detection in a method based on allele-specific primer extension (Figure 1.5B) called allele-specific PCR (AS-PCR), also referred to as the amplification refractory mutation system (ARMS; Newton *et al.,* 1989), allele-specific amplification (ASA), or preferential amplification of specific alleles (PASA; Sommer *et al.,* 1992). In conventional AS-PCR, two PCR primers are designed with the 3'-terminal nucleotide complementary to either nucleotide of a SNP (wild type or mutant), used in combination with a common reverse primer to selectively amplify the SNP alleles in two separate PCR tests (Newton *et al.,* 1989). By using a *Taq* DNA polymerase lacking 3'→5' proof-reading activity, specificity is maintained since, under optimised conditions, *Taq* DNA polymerase will not extend a primer with a 3' mismatch. Therefore, a single product will indicate homozygous inheritance of a particular allele, whilst a product in both reactions indicates heterozygosity. An internal control set of primers is used to amplify a region spatially separated from the target region, ensuring that a negative result is attributable to the absence of a specific allele, and not a failure of the PCR.

The traditional approach used for analysing AS-PCR products is by electrophoretic size separation of DNA fragments on a gel matrix, followed by staining with a fluorescent double-strand DNA (dsDNA) specific binding dye such as ethidium bromide (Markovits *et al.,* 1979; Figure 1.6A). Ethidium bromide (2,7-diamino-10-ethyl-9-phenylphenanthridinium bromide) has a planar structure allowing intercalation into approximately every second base pair of dsDNA.

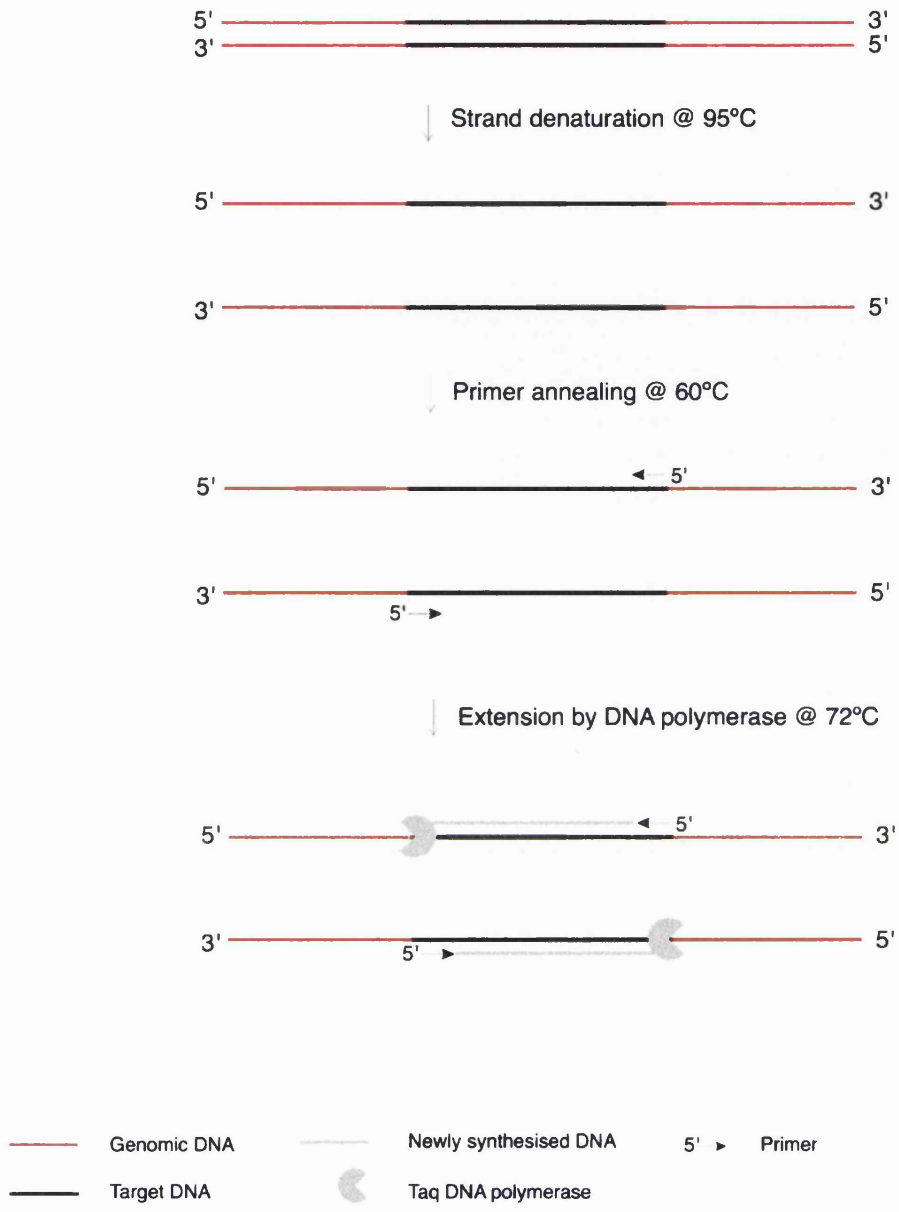


Figure 1.7: A schematic representation of PCR dynamics, showing the three temperature stages of a PCR cycle, including denaturation, annealing and extension.

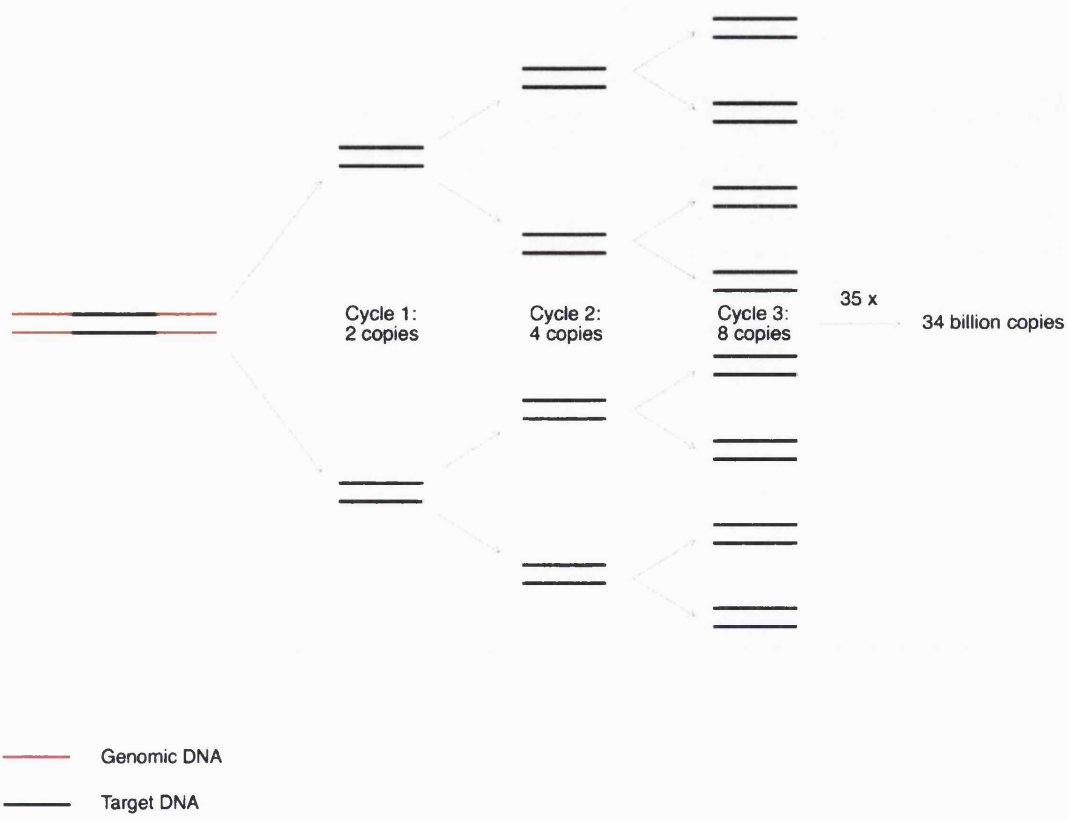


Figure 1.8: Simplified representation of the exponential amplification of nucleic acid during a PCR test.

When ethidium bromide is bound to DNA and excited by UV light, it exhibits a greater fluorescence enhancement than when free in solution, allowing ethidium bromide: DNA complexes to be readily visualised (Sharp *et al.,* 1973). The fluorescent enhancement is thought to result mainly from two factors: immobilization of the dye molecule in a planar conformation (allowing extensive electron delocalisation to occur) and exclusion of solvent (water) from interactions with the dye that might otherwise quench fluorescence (Mason, 1999; Cosa *et al.,* 2001).

In 1992, a pioneering application of ethidium bromide was published in which the dye was incorporated directly into a PCR reaction mixture to visualise DNA amplification in real time by directing excitation illumination through walls of the amplification vessel before or after, or even continuously, during thermal cycling (Higuchi *et al.,* 1992). This work was pivotal in revealing the shape and characteristics of a PCR reaction curve (Figure 1.9), prompting the development of PCR instruments with integrated optics systems allowing continuous monitoring of DNA amplification in a closed reaction vessel.

This advancement was in line with the synthesis of new dyes (unsymmetrical cyanine dyes) including SYBR® Green I, YOYO®, YO-PRO®, and SYBR® Gold (Tuma *et al.,* 1999), which exhibit more favourable properties for DNA analysis than ethidium bromide, principally because they are less mutagenic (Lunn and Sansone, 1987). As a group, the unsymmetrical cyanine dyes are characterised as having low intrinsic fluorescence, large fluorescent enhancements upon binding to nucleic acids, and high fluorescence quantum yields when complexed with nucleic acids. These features combine to produce excellent signal-to-background noise ratios. In combination with their apparently high binding affinities for nucleic acids, and greater sensitivity than ethidium bromide, these dyes are a favourable option for inclusion in PCR reaction mixes for real-time monitoring of PCR (Lay and Wittwer, 1997) and quantitative PCR strategies (Rasmussen *et al.,* 1998; Section 1.3). Although the mechanism of binding is still unknown for these dyes, it is thought to bind to the minor groove of the DNA duplex (Singer *et al.,* 1997).

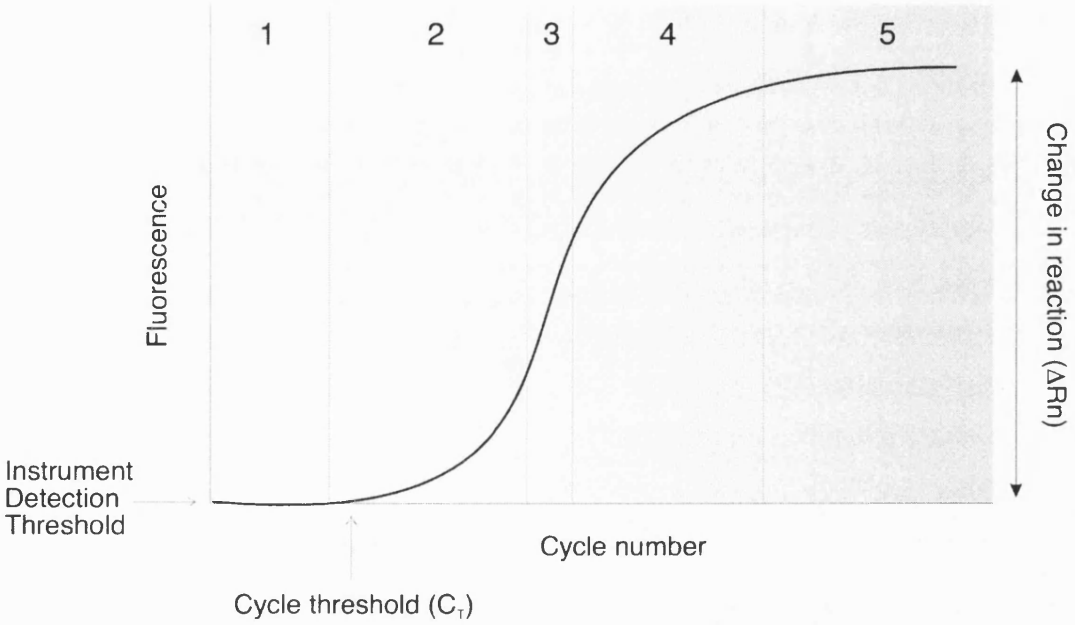


Figure 1.9: The five phases of a typical real-time PCR curve. *Phase 1-* Baseline, where increases in fluorescence are below noise level. Often this is used to derive the 'cycle threshold' value, defined as the fractional cycle number at which the fluorescence passes a fixed threshold. A fixed threshold can also be set above the baseline. *Phase 2-* Log-linear or exponential phase, where the reaction proceeds at maximum efficiency displaying an exponential increase in product accumulation (doubling per cycle). *Phase 3-* linear amplification phase, where the reaction no longer exhibits an exponential amplification, rather a linear increase in product. *Phase 4-* exponential decay, where the reaction efficiency is decreased and a slower production of product ensue. *Phase 5-* plateau phase where there is no detectable increase in amplification, possibly due to high substrate inhibition (Kainz *et al.*, 2001) or a reduction in the availability of critical reaction components.

Double-strand specific DNA binding dyes also enable the real-time monitoring of a methodology called DNA melting curve analysis, which allows product amplification *and* identification in the same reaction vessel. 'DNA melting' describes the denaturation of DNA from a double-stranded to single-stranded state. The specific temperature at which the DNA is 50% single stranded and 50% double stranded is known as the "DNA melting temperature" (T_m; Ririe *et al.,* 1997). Conventionally, DNA melting is monitored by changes in optical density as a sample temperature is slowly elevated. The principle behind this is the ability of the heterocyclic rings of the bases to adsorb light strongly in the ultraviolet range (with a maximum close to 260 nm). When DNA is in its double helical state the electron systems of the heterocyclic rings overlap, made possible by their stacking in the parallel array of the double helix. This reduces their ability to absorb UV light. During DNA melting, the electron systems are freed from the constraints of base pairing and have the ability to absorb UV light. This is known as the 'hyperchromic' effect. Any departure from the duplex state of DNA is immediately reflected by a decline in the hyperchromic effect, exhibiting an increase in optical density toward the characteristic of free bases (Chamberlin and Berg, 1963).

Continuous fluorescence monitoring of a PCR mixture containing DNA and a dsDNA specific dye, whilst slowly elevating the temperature shows a rapid loss of fluorescence at a particular temperature where the helical dsDNA is denatured to single strands and the dye can no longer bind. A plot of fluorescence vs. temperature is called a "melt curve" (Figure 1.10A), which is converted to a "melt peak" by plotting the first negative derivative of fluorescence loss, F, with respect to temperature, T ($-dF/dT$; Figure 1.10B). Melt peaks allow easy visual interpretation of melt curve data, allowing the identification of a PCR product (Ririe *et al.,* 1997). The shape and position of the melt curve is mainly influenced by base composition, length and sequence of DNA strand (i.e. GC/AT ratio), allowing products of a similar size to be differentiated in a reaction mixture (Ririe *et al.,* 1997). Although the dynamic range of discrimination using melt curve analysis is limited, greater product differentiation can be achieved using smaller amplicons (50-150 bp). More-refined temperature measurement by instrumentation may allow a wider discrimination of T_m (Ririe *et al.,* 1997).

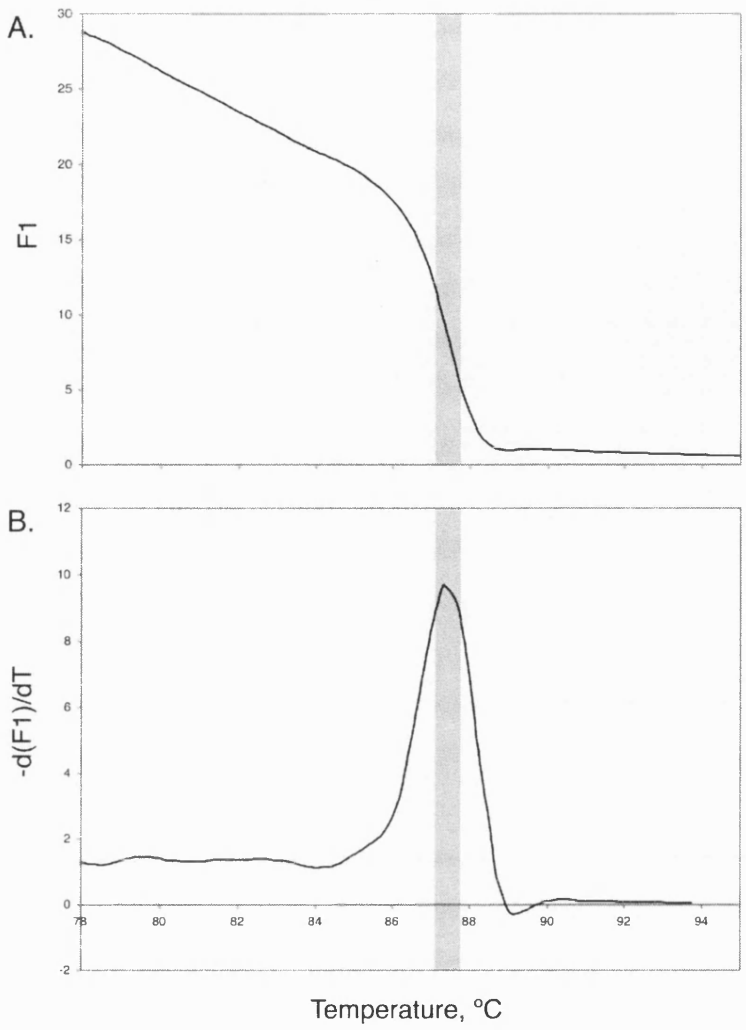


Figure 1.10: [A] Typical melt curve in which the sample is slowly heated to complete denaturation. The rapid loss of fluorescence occurs at the specific melting temperature of the PCR amplicon where the dsDNA binding dyes can no longer bind. [B] Typical melt peak, where the first negative derivative of the fluorescence (F1) data in [A] with respect to the temperature, T, is plotted versus temperature. Grey band demonstrates where the most rapid loss of fluorescence occurs, translating into the melt peak by differentiation. In addition to factors described in text, position of melt peak can be influenced by ionic strength (Schildkraut and Lifson, 1965), concentration of fluorescent dye, number of products in reaction mixture, and temperature gradient used to perform melt analysis (Ririe *et al.,* 1997).

The information obtained from a melt curve analysis is not strictly "sequence specific" and therefore artefactual amplification products can interfere with the data. Often these artefacts melt at a lower temperature than a specific product, and show a broader melt peak allowing their differentiation. Even so, using SYBR® Green I to monitor PCR reactions remains a popular choice among researchers, and has been used for a range of novel assays including McSNP® genotyping (Figure 1.5F; Akey *et al.,* 2001), and homogenous allele-specific primer extension tests (Table 1.1).

An established alternative to allele-specific primer extension methods for distinguishing between SNP alleles is to monitor a PCR reaction by allele-specific hybridisation (Figure 1.5A) of real-time PCR probes. Critically, these probes employ hybridisation of the target nucleic acid sequence to its complement; unlike dsDNA binding dyes this adds sequence specificity to genotyping assays. The assays are based on the same biochemical principle as the original allele-specific hybridisation publications introduced earlier (Wallace *et al.,* 1979; Chang and Kan, 1982), but the advancements in reagents, protocols and instrumentation have contributed to the progression of hybridisation formats. Instead of using lengthy dot-blot assays the hybridisation event is usually reported by FRET mechanisms (Figure 1.6D).

FRET is a distance-dependent interaction between the electronic excited states of two dye molecules in which energy is transferred from a donor molecule to an acceptor molecule without emission of a photon. For efficient FRET, there must be sufficient spectral overlap between donor emission and acceptor absorption wavelengths (Figure 1.11). Also, donor and acceptor transition dipole orientations must be approximately parallel (Stryer and Haughland, 1967). The efficiency of energy transfer (E) is defined as the number of quanta transferred to the acceptor, divided by the number of quanta absorbed by the donor. It is described by the expression:

$$E = (R_0/R)^6 / 1 + (R_0/R)^6$$

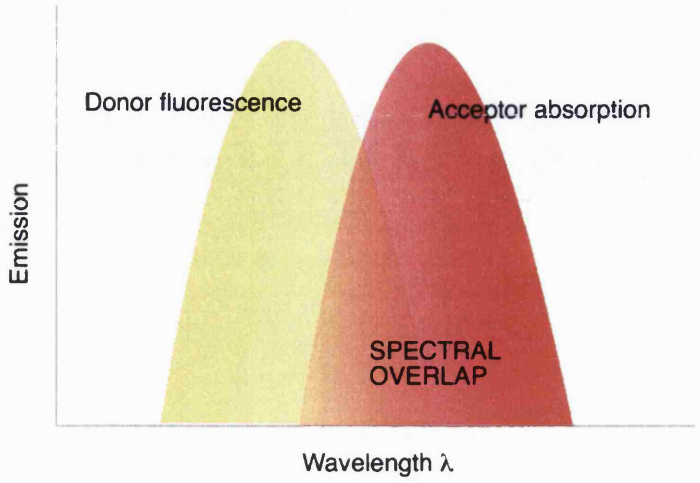


Figure 1.11: A specific requirement for FRET to occur is that there is sufficient spectral overlap between the donor fluorescence wavelength, and the acceptor absorption wavelength.

Primer extension	*Based on the selective extension of oligonucleotides by Taq DNA polymerase*				
Reference	**Method**	**Concept**	**Advantages**	**Disadvantages**	**PCR step**
Newton et al., 1999	Gel electrophoresis	Allele-specific PCR performed in solution phase, and products identified by electrophoresis	Inexpensive	Time consuming, multiple steps	Yes
Germer and Higuchi, 1999	Intercalating dyes	Allele-specific PCR performed in a single tube, one allele-specific primers has a GC tail to increase the T_m of the specific product.	Rapid, homogenous, inexpensive	Minimal multiplexing capacity	Yes
Myakishev et al., 2001	FRET primers	Allele-specific PCR probes used to report a mismatch by maintaining FRET	Simple, allows multiplexing	Expensive probes, complex design rules	Yes
Waterfall and Cobb, 2002	Intercalating dyes	Single tube adaptation of allele-specific PCR where primers are arranged in bi-directional manner around the SNP. Products identified by different T_ms.	Rapid, homogenous, inexpensive	Minimal multiplexing capacity	Yes

Table 1.1: Common genotyping assays based on allele-specific primer extension. The list is not intended to be extensive due to the number of similar assays, but rather provide details of a range of assays.

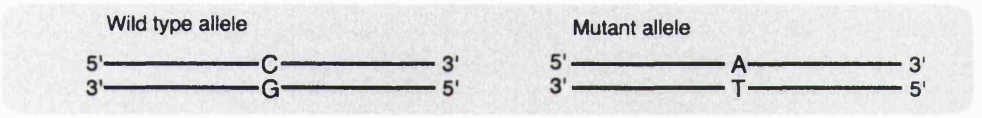
where R is the distance between the donor and acceptor (in Å) and R_O (in Å), the distance at which energy transfer is 50% efficient, also called the Förster radius (Förster, 1946; Förster, 1948).

This formula shows that the rate of energy transfer is inversely proportional to the sixth power of the distance between the donor and acceptor fluorophores. Therefore, the efficiency of the transfer rapidly declines to zero at distances larger than the Förster radius. The Förster radii have to be experimentally measured for each specific donor-acceptor pair, but generally for FRET to occur, the donor and acceptor molecules have to be within 1-10 Å of each other.

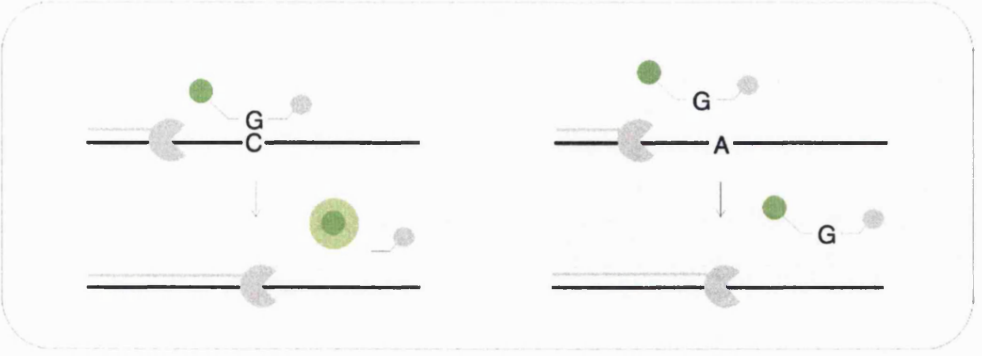
If the acceptor is a fluorescent dye itself, the transferred energy can be emitted as a fluorescence characteristic of the acceptor. If the acceptor is not fluorescent, the energy is 'quenched' through equilibration with solvent, and the acceptor is then known as a 'quencher'. A quencher is a molecule that absorbs or dissipates energy from an excited fluorophore (the reporter), returning it to the ground state without any fluorescent emission by that fluorophore (Didenko, 2001).

Common probe mechanisms utilising FRET for real time PCR monitoring and SNP detection are summarised in Figure 1.12. The most popular probe configuration for real-time PCR monitoring is the 5' fluorogenic nuclease probe (commercially known as TaqMan® probe; Figure 1.12A). A TaqMan® probe is an oligonucleotide labelled with a reporter dye and a quencher dye, protected from extension by *Taq* DNA polymerase by phosphorylation at the 3' end (Livak, 1999). During the extension phase of a typical PCR run, the 5' nuclease activity of *Taq* DNA polymerase hydrolyses the probe in the 5' to 3' direction, liberating the reporter dye from the quencher dye, disabling FRET. A cumulative increase in reporter fluorescence indicates that the probe-specific target has been amplified (Holland *et al.,* 1991). TaqMan® probes represent a category called 'hydrolysis' probes, named as such because the probe is hydrolysed to release the reporter dye (consumed by the 5' nuclease activity of DNA polymerase).

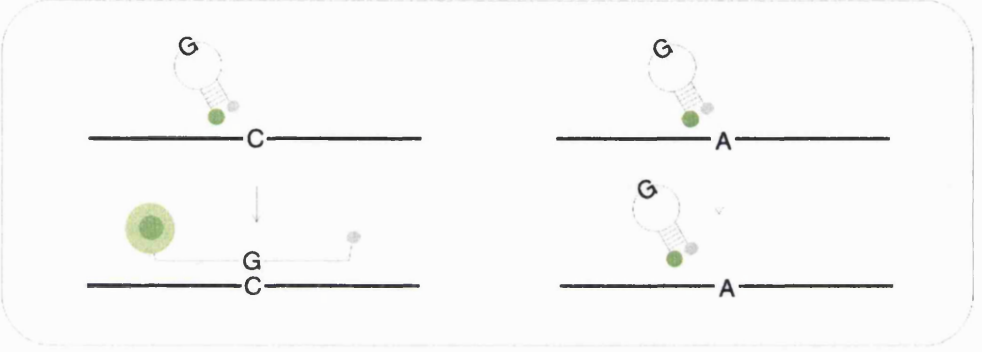
Figure 1.12: Principles for homogenous SNP genotyping by real-time PCR. This list is not intended to be comprehensive. The grey panel demonstrates a C-to-A polymorphism. Probes shown are designed to detect the wild-type allele. [A] *Fluorogenic 5' nuclease (TaqMan®) probes:* Two allele-specific oligonucleotide probes are labelled with two fluorescence reporter dyes emitting at different detectable wavelengths. Probes blocked from extension by Taq by a 3' phosphorylation. Hydrolysis of the probe by the 5' nuclease activity of *Taq* only occurs if the probe is specifically hybridised to a perfectly matched target sequence. Liberation of the reporter dye from the quencher prevents FRET, providing a fluorescent signal characteristic of the specific probe (Livak, 1999). [B] *Molecular beacons:* Two allele-specific probes, labelled with fluorophores emitting at different detectable wavelengths, are designed such that a perfectly matched probe-target hybrid will be energetically more favourable than the stem-loop structure. A mismatched probe-target hybrid favours the stem-loop conformation. A perfectly matched sequence enables hybridisation to the DNA sequence, preventing FRET quenching, and emitting light at a characteristic wavelength (Tyagi *et al.,* 1998) [C] *Dual hybridisation probes:* Two probes are included, one is 3'-labelled with either a donor or acceptor dye, and the other is 5'-labelled with either a donor or reporter (acceptor) dye. Probes are PCR-blocked at the 3' end by phosphorylation. Perfectly matched probes are in close enough proximity enabling FRET, giving a signal characteristic of the acceptor dye. A mismatch causes one of the probes to melt off at a detectably lower temperature (De Silva *et al.,* 1998). [D] *Universal Energy-Transfer Labelled probes (Amplifluor®):* An allele-specific primer is designed with a 5' tail sequence (Z) that is not complementary to the target. Probe consists of a molecular beacon with no specificity for the target product, attached to a primer with a sequence (Z') complementary to the Z sequence of the allele-specific primer. The area between the primer and beacon is not PCR-blocked. During amplification, the complement sequence of the beacon is amplified allowing product and probe to anneal. This opens the beacon structure, preventing FRET quenching. Using beacons with different Z' sequences and allele-specific primers with different Z sequences allows product identification by characteristic fluorescence emissions (Myakishev *et al.,* 2001). [E] *Scorpion primer:* Two allele-specific probes are labelled with fluorescent dyes emitting at different detectable wavelengths. Variant base can be designed at the 3' end of the primer portion of the probe, using allele-specific PCR as the discrimination mechanism (as figure), or by detecting a mismatch within the stem loop portion as [B]. After extension, during PCR, the specific probe sequence binds to its complement within the same strand of DNA, opening the stem-loop and disabling FRET. An increase in signal is observed. In the allele-specific probe embodiment, a mismatch prevents amplification of complementary sequence, for loop to bind to. If the mutation is in the stem-loop portion, a mismatch will cause FRET to remain active by favouring the closed conformation (Whitcombe *et al.,* 1999). [F] *HyBeacon® probes:* Oligonucleotide probes complementary to the wild type or mutant sequence, internally labelled with a specific fluorophore. Probes are blocked from PCR extension at the 3' end by phosphorylation. Increased florescence is observed when the probe is bound to a perfectly matched sequence (French *et al.,* 2001). [G] *Resonsense® probes*: An oligonucleotide probe is labelled with a single fluorophore and blocked at the 3' end from extension. An exonuclease deficient *Taq* is used to prevent hydrolysis during the PCR. Probe is designed to hybridise to a perfectly matched target sequence generating a product probe duplex for a DNA intercalater to bind (SYBR® Gold). The intercalater is used as the donor moiety to fluoresce when bound to dsDNA and excite the fluorophore on the probe by FRET. [H] *Angler® probes:* Similar to [G] except that the probe is chemically linked to the primer bringing it spatially close to the extended primer for sequence specific self-probing of product. In this case, the 3' end of the primer is not protected from extension (Lee *et al.,* 2001). [I] PNA-based Light-Up probes: Probes are made up of peptide nucleic acids labelled at the amino-terminus with a Thiazole orange moiety. When the probe is bound to perfectly matched DNA, the dye fluoresces (Wolffs *et al.,* 2001).



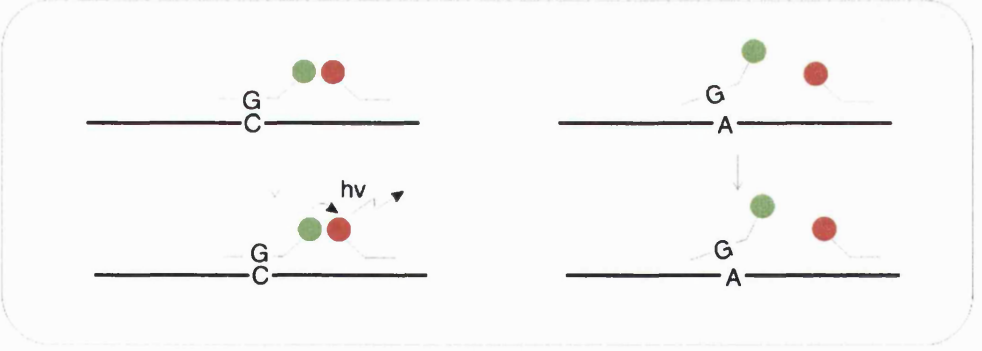
A. 5' nuclease (TaqMan) probes

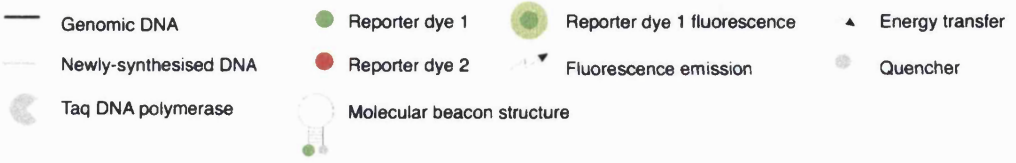


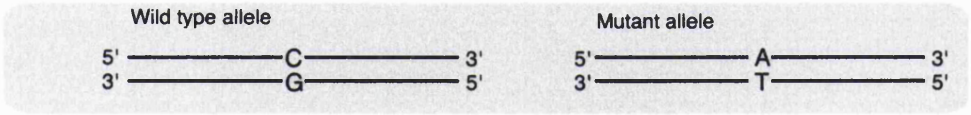
B. Molecular beacons



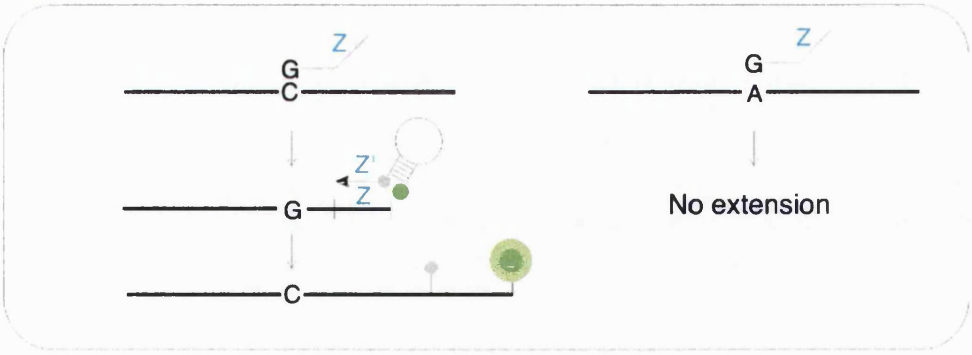
C. Dual hybridisation probes



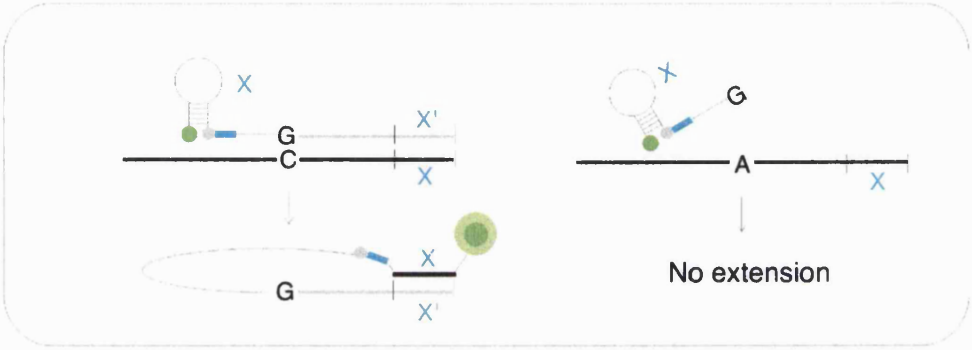




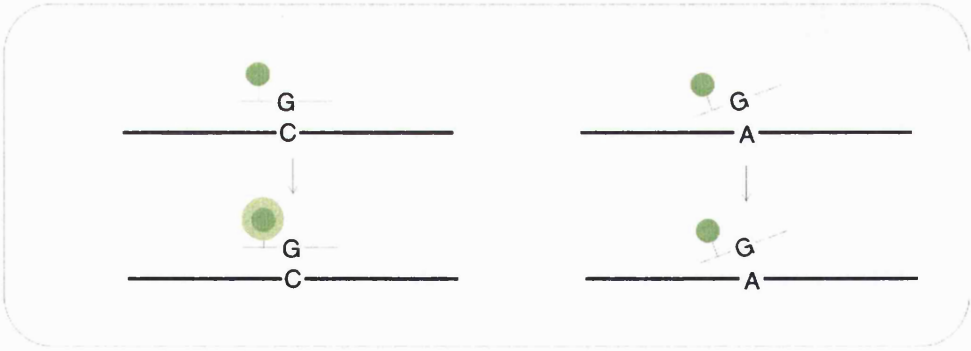
D. Universal FRET probes



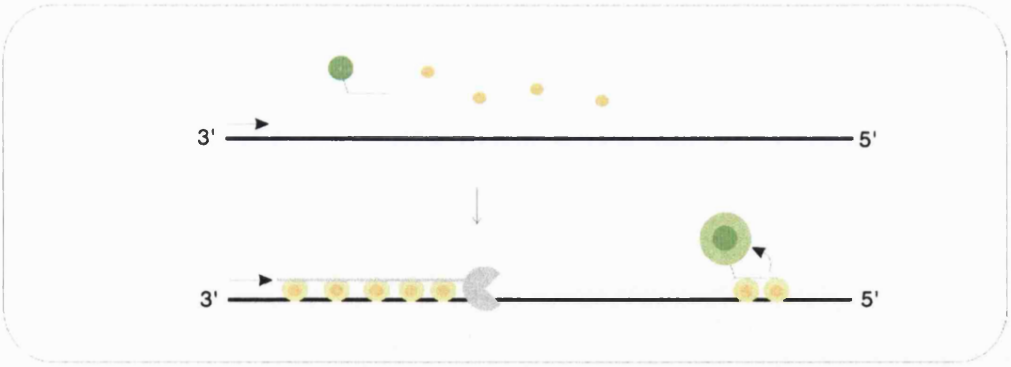
E. Scorpion primers



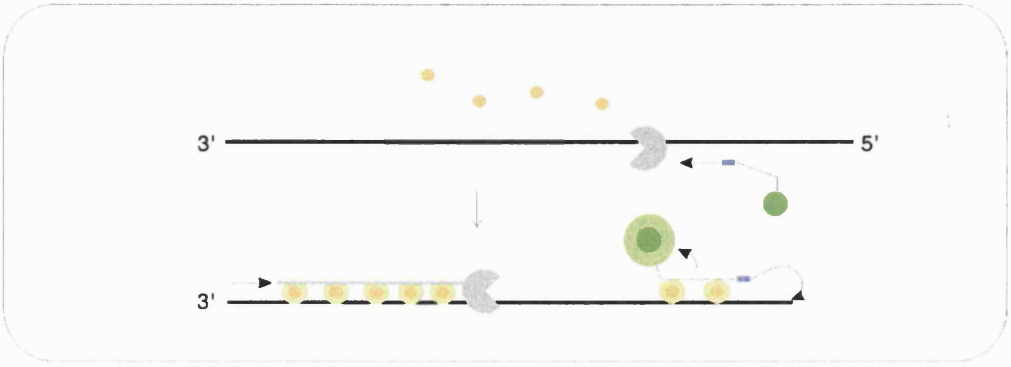
F. Hybeacon probes



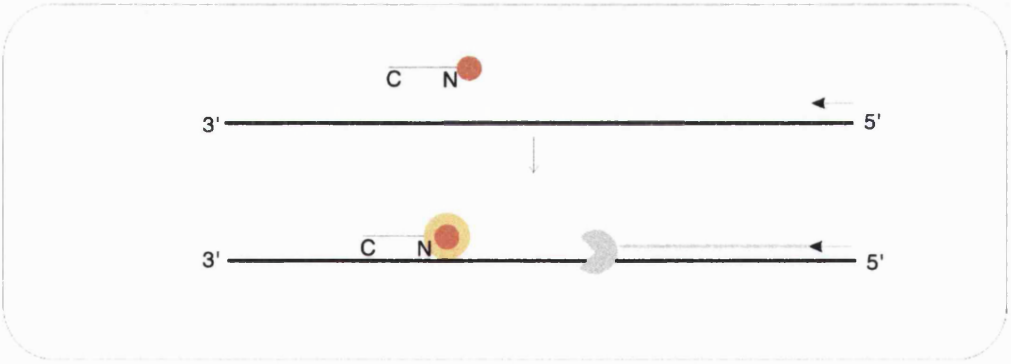
G. Resonsense probes



H. Angler probe



I. Light-up probes



The hybridisation aspect of TaqMan® probes is apparent during SNP detection where cleavage of the probe only occurs if the probe is specifically hybridised to its target sequence (Livak, 1999). By including two probes complementary to each allele labelled with different fluorescence reporter dyes at the 5' end with detectably different emission wavelengths (usually FAM, 6-carboxy-fluorescein, λmax 535 nm, and JOE, 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein, λmax 557 nm) into a single PCR test, only the matched probe will produce a significant measured signal increase because a matched conformation provides the desired substrate for 5' nuclease activity (Figure 1.13). A mismatch between the probe and target greatly reduces hybridisation and cleavage efficiency because this specific recognition structure is disrupted.

Although the TaqMan® assay is an established technique for genotyping, conditions that distinguish between probes with a single nucleotide difference can be difficult to obtain (Sevall, 2000). A necessary requirement is that the TaqMan® probe is bound to the template at temperatures where *Taq* polymerase is optimally extending primers (60-72°C). This usually requires an oligonucleotide length of 25-40 nucleotides which is generally too long for most SNP detections given that the difference in T_m between a perfectly matched duplex 25 bp long and a similar duplex with a single mismatch is usually only about 3°C (Walburger *et al.,* 2001). Without careful optimisation of probe sequence and PCR conditions it can be difficult to detect SNPs with an acceptable signal-to-noise ratio for a robust clinical assay.

The remaining probe formats described in Figure 1.12 report changes in the proximity of reporter and quencher dyes by FRET directed by a hybridisation event, not hydrolysis. During each cycle, the probes hybridise to the target sequence present to give a signal that corresponds to the amount of PCR product in each cycle, as opposed to the cumulative signal achieved with TaqMan® (hydrolysis) probes.

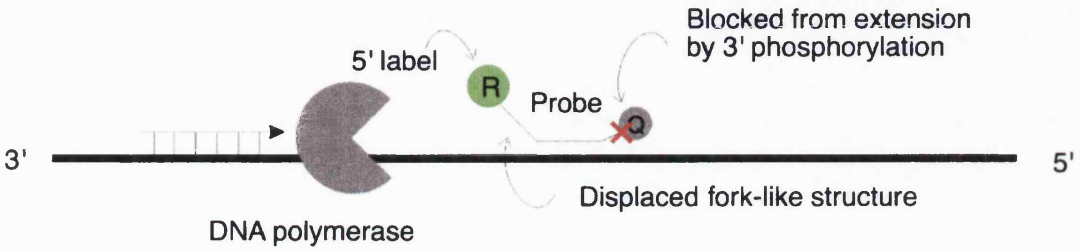


Figure 1.13: The substrate for *Taq* DNA polymerase in TaqMan® probe assays. This substrate is a forked structure occurring at the phosphodiester bond joining the displaced region with the base-paired portion of the strand (Figure 1.13; Holland *et al.*, 1991).

For dual hybridisation probes (Figure 1.12C) the perfect hybridisation of two labelled probes enables FRET to occur. Following excitation of the fluorophore attached to the donor probe by a light source, the fluorophore attached to the reporter probe can be excited by FRET mechanisms, generating a signal that is characteristic of the reporter dye. As described in Figure 1.11, the emission wavelength of the donor dye must have sufficient spectral overlap with the excitation wavelength of the reporter dye for efficient FRET to occur. It is the reduced hybridisation stability of a mismatched probe, which causes it to melt off at a detectably lower temperature than a match (disabling FRET excitation of the acceptor probe) that enables efficient SNP detection. A similar concept is employed in Resonsense® and Angler® probe formats (Figure 1.12G and 1.12H, respectively; Lee *et al.,* 2001) where the fluorescence emission of a dsDNA specific binding dye (SYBR® Gold) is used to excite a sequence-specific reporter probe, as opposed to an adjacent sequence-specific donor probe.

A popular configuration used in the design of hybridisation probes is a stem-loop structure (or hairpin), commercially referred to as a molecular beacon (Figure 1.12B). Molecular beacons are stem-loop structures labelled at the end of the 5' arm with a reporter dye, with a quencher dye attached to the end of the 3' arm. When free in solution, the molecular beacon adopts the stem-loop structure that brings the reporter dye and quencher into close proximity, allowing FRET quenching. In the presence of a complementary sequence, hybridisation to the target is thermodynamically more favourable than the closed stem-loop configuration, which increases the distance between reporter and quencher dyes such that efficient quenching by FRET cannot occur. A fluorescent signal characteristic of the reporter dye ensues. It follows that in the presence of a mismatched sequence, the stem-loop structure is energetically favoured and quenching by FRET is maintained (Tyagi *et al.,* 1998).

Certain probe systems have combined the molecular beacon structure with allele-specific primer extension to introduce 'semi' sequence specificity to AS-PCR methods. One example, the Amplifluor™ system, involves PCR amplification of genomic DNA with two allele-specific primers that have a 5' tail sequence attached.

The specific tail sequences contain primer-binding sites for two differentially labelled molecular beacon primers called 'universal energy-transfer-labelled primers'. In the presence of a matched sequence, the appropriate allele-specific primer will be extended, and the specific 5'-tail sequence becomes part of the PCR product. Therefore, the corresponding universal energy-transfer-labelled primer can hybridise to the PCR product, opening the stem-loop structure, preventing FRET quenching. A signal characteristic of the reporter dye will ensue. It follows that *Taq* will not extend the mismatched allele-specific primer, and therefore the complementary 5' tail will not be amplified. With different 5' tail sequences, semi-sequence specificity is introduced into the assay (Figure 1.12D; Myakishev *et al.,* 2001).

A further example of combining molecular beacons and allele-specific PCR is scorpion primers. Here, a molecular beacon is directly attached to an allele-specific oligonucleotide (Figure 1.12E). The beacon is designed with complementarities to the sequence upstream of the SNP site. In the presence of a matched sequence, the allele-specific oligonucleotide portion of the probe is extended, and a complement to the stem-loop structure is produced in the PCR product. Consequently, the beacon portion will hybridise to this sequence, disabling FRET quenching and a characteristic signal ensues. A mismatch at the 3' end of the allele-specific portion of the probe will prevent extension of the primer, and the complementary sequence for probe hybridisation is not formed. Therefore, the scorpion maintains its stem-loop structure without generating a signal (Whitcombe *et al.,* 1999). Scorpion probes have also been designed with the base variant occurring in the stem-loop portion of the probe, exploiting the thermodynamic properties of this structure, as described for molecular beacons (Thelwell *et al.,* 2000).

More recently, probe systems have been described that rely on alterations in detectable fluorescence caused by hybridisation of a *singly* labelled probe to their complement. These include HyBeacons® (French *et al.,* 2001; Figure 1.12F), 5'-Labelled Oligonucleotide probes (Crockett and Wittwer, 2001) and Light Up® probes (Wolffs *et al.,* 2001; Figure 1.12I). For HyBeacons®, a single fluorophore is attached

to an internal 'T' residue of an oligonucleotide probe. An elevation in fluorescence is observed on hybridisation to its complimentary sequence within the PCR amplicon. Here, the proposed mechanism stems from fluorescence quenching by the nucleotide residues in the probe (in particular purine bases) when in solution; this is likely to be a result of base-dye stacking and electron transfer events (French *et al.,* 2001). This is analogous to Light Up® probes except the oligonucleotide is replaced by a peptide nucleic acid (PNA) backbone, which has an extremely high affinity for DNA (Nielsen *et al.,* 1991). Conversely, the 5'-labelled oligonucleotide probe (Crockett and Wittwer, 2001), exhibits a fluorescence quenching activity when bound to its complementary target.

All three single-labelled probe systems rely on the inherent quenching characteristics of nucleotides themselves, resulting in different alterations in measurable fluorescence. The advantage of these probes is that they contain a single fluorophore lowering the overall cost of reagents and synthesis. However, the effectiveness of the probe is highly influenced by the surrounding or complementary sequence. HyBeacons® require an internal 'T' nucleotide on which to attach the fluorophore, and the 5' labelled probe exhibits greater effects during hybridisation to 'G'-rich amplicons (Crockett and Wittwer, 2001). This may not be possible in all sequences containing SNPs.

Probe-based systems have increased the multiplexing opportunities for SNP detection due to the commercial availability of a huge variety of fluorophores with detectable differences in their emission wavelengths. These probes can be rationally designed using specific rules to recognize and detect any desired target with almost any desired degree of specificity by varying length, sequence, and hybridisation conditions. However, often the probe design rules can be complex and prohibiting, requiring specialised software (Täpp *et al.,* 2000). In addition, the actual cost of real-time PCR probes is high in comparison with simple intercalating dyes. With the increased commercial availability of novel probes, these costs could be driven down by virtue of competition.

So far, the use of allele-specific probes in a solution phase has been described. Alternatively, probes can be immobilised on a solid-support in high throughput microarray formats. Microarrays vary according to the solid support used (such as glass or filters), the surface modifications with various substrates, the type of nucleic acid fragments on the array (such as cDNA, oligonucleotides or genomic fragments), whether the gene fragments are pre-synthesised and deposited, or synthesised *in situ*, and the machinery used to place the fragments on the array (such as ink-jet printing, spotting, or *in situ* synthesis; Holloway *et al.,* 2002). However, the basic concept behind all microarrays is the precise positioning of the nucleic acid probes onto a solid support such that they can act as molecular detectors. The specific hybridisation of nucleic acid from a single sample, to thousands of these molecular detectors (or probes), can be monitored using a range of different methods (Figure 1.6), generating a great deal of information from a single sample simultaneously.

For SNP detection, the thermal stability of a hybrid formed between an allele-specific oligonucleotide (ASO) probe and its corresponding SNP-containing target sequence is determined by the stringency of the reaction conditions and the surrounding sequence of the SNP. Therefore, it is very difficult to establish conditions that are optimal for genotyping all SNPs. This limits the multiplexing capabilities. Using nucleotide analogues with a higher affinity for DNA than DNA itself, e.g. locked nucleic acids (LNAs) or peptide nucleic acids (PNAs; Figure 1.14), more robust genotyping can be achieved using shorter probes than natural ASO probes since a mismatch in either of these chemistries causes a greater disruption of hybridisation than a simple DNA-DNA mismatch (Orum *et al.,* 1999; Jensen *et al.,* 2001). Details of other allele-specific hybridisation methods are summarised in Table 1.2.

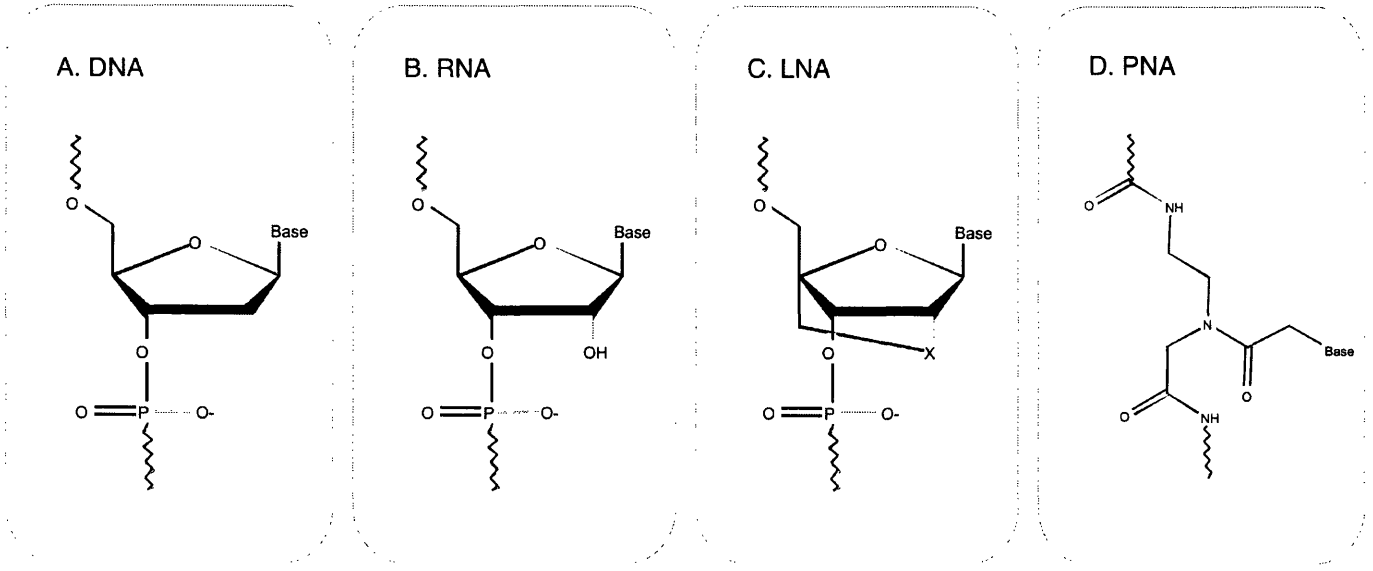


Figure 1.14: [A] DNA monomer, [B] RNA monomer, [C] Locked Nucleic Acid (LNA) monomer, and [D] Peptide Nucleic Acid monomer. The LNA monomer is a bicyclic compound in which the 2' and 4' positions of the furanose ring are linked by an O-methylene (oxy-LNA), S-methylene (thio-LNA) or NH-methylene (amino-LNA) moiety, represented by X in [C]. This linkage restricts the conformational freedom of the furanose ring, conferring high thermal stability when hybridised to a complementary DNA (or RNA) sequence (Koskin *et al.*, 1998, Singh *et al.*, 1998; Orum *et al.*, 1999). The PNA polynucleotide has a 2-aminoethylglycine backbone in lieu of the deoxyribose phosphate backbone of DNA. Like LNAs, PNA-DNA duplexes are also more sensitive to mismatches than DNA-DNA mismatches (Orum *et al.*, 1993).

Hybridisation	*Based on the thermal stability of a hybrid formed by an allele-specific probe and a SNP-containing target sequence*				
Reference	**Method**	**Concept**	**Advantages**	**Disadvantages**	**PCR Step**
Livak, 1999	TaqMan probes (ABI)	Solution phase, homogenous, real-time PCR assay. Probe hybridises between primer annealing sites.	Simple assay format uses 5' nuclease activity of Taq to digest a fully matched probe, disabling FRET quenching	Probes are expensive, complex design considerations	Yes
Tyagi et al., 1998	Molecular Beacons	As above, except probe has a hairpin stem-loop structure.	Simple assay. Matched probe hybridisation disables FRET quenching.	Probes are expensive, complex design considerations	Yes
Orum et al., 1999	LNA probes (Genset)	A replacement for phosphodiester oligonucleotide probes attached to an array to capture complementary sequences.	More specific than phosphodiester probes, shorter than oligonucleotide probes.	Limited availability, slightly more expensive than simple phosphodiester oligonucleotide probes.	Yes
Saiki et al., 1989	Reverse dot blot	Based on hybridisation between ASO probes that have been immobilised on a membrane, and PCR products in solution.	One of the original genotyping methods used, multiplexing possible	Prone to low signal-to-noise ratio, time consuming	Yes
Lipshutz et al., 1985; Mei et al., 2000	GeneChip (Affymetrix)	Solid support genotyping on a high density chip	Very high probe density, excellent multiplexing capabilities	High failure rates, expensive for design and manufacture of chip, set of markers cannot be changed quickly or arbitrarily, long protocol	Yes
Howell et al., 1999	DASH (Dynametrix Ltd.)	Typically a solid support assay, looks for the melting temperature differences of allele-specific probe when hybridised to a matched or mismatched target	Inexpensive labelling using intercalater dyes, uses biotinylated PCR product to maintain single stranded nature for capturing	Complex design rules, specialised instrumentation required, single reporter means two reactions must be run in parallel per SNP	Yes

Table 1.2: Common genotyping assays based on allele-specific hybridisation. The list is not intended to be exhaustive due to the number of similar assays, but rather provide details of a range of assays.

A further principle in which the specific properties of DNA polymerases are used is in allele-specific nucleotide incorporation (Figure 1.5C; Table 1.3), or single base extension (minisequencing) methods. These are becoming the reaction principle of choice for high throughput SNP genotyping projects because of the high accuracy of nucleotide incorporation by DNA polymerases. The primer extension reaction is robust, and allows genotyping of most SNPs without a large variation in reaction conditions. These features are advantageous for high throughput projects involving SNP validation, because the effort required for assay design and optimisation are minimised. A range of detection methods is available for single base extension assays, including mass spectrometry (Figure 1.6 B), FP (Figure 1.6 C) and FRET (Figure 1.6 D). Generally, a PCR-step is required to achieve sufficient target sequence for SNP detection.

DNA ligases have also been in genotyping assays, based on the ability of the enzyme to discriminate against mismatches at the ligation site in two adjacently hybridised oligonucleotides (Figure 1.5D; Table 1.4). This reaction principle has been described with multiplex detection using fluorescently labelled oligonucleotide probes in the template dye-incorporation assay (TDI, Figure 1.15; Chen *et al.,* 1998), and combined with colourimetric detection of hybridised probes in microtiter plates (Lizardi *et al.,* 1998). Ligase-mediated genotyping is also used in the dye-labelled oligonucleotide ligation (DOL) assay, with detection by FRET or FP (Figure 1.16; Chen *et al.*, 1999).

The genotyping methodologies introduced during this chapter have all required a specific amplification step, which is invariably PCR, to achieve the required sensitivity. In addition, PCR allows accurate DNA quantification (Section 1.3), a procedure that is highly valuable to clinical laboratories assaying viral loads, gene copy numbers or bacteria concentration in human specimens. The progress in PCR instrumentation has also allowed the development of homogenous assays that combine amplification and detection in a single tube, offering huge advantages to clinical laboratories, where separate steps are to be avoided to minimise sample mix-ups and contamination.

Nucleotide incorporation	*The distinction between genotypes is based on the high accuracy of nucleotide incorporation by Taq DNA polymerase*				
Reference	**Method**	**Concept**	**Advantages**	**Disadvantages**	**PCR step**
Nikiforov et al., 1994	SNPstream (Orchid), GBA	Incorporated hapten-labelled nucleotide analogues recognised by antibodies. Enzymes conjugated to antibodies catalyse formation of a coloured product	Inexpensive, robust, automated HTP process (384-well)	Multiple detection steps	Yes
Dubiley et al., 1999	Multiplex minisequencing SnaPShot (ABI)	Fluorescently labelled dideoxynucleotides used in minisequencing reactions.	Multiplexing capacity by using primers modified with 5' tails of varying lengths, compatible with capillary DNA sequencers.	Size separation step	Yes
Ronaghi et al., 1996	Pyrosequencing (Pyrosequencing AB)	Inorganic pyrophosphate (PPi) released from a nucleoside triphosphate during DNA chain elongation, detected by a luminescence assay	Sequencing of up to 50 bases, operates in 96-well or 384-well format	Expensive, difficult to multiplex, dedicated instrument	Yes
Ross et al., 1998	MassEXTEND, MassARRAY (Sequenom)	Primers of different lengths used with mixtures of dNTPs and ddNTPs yield allele-specific primer extension products with detectable differences in molecular mass.	High multiplexing capacity, high resolution, uses MALDI-TOF to detect the primer extension products.	Expensive instrumentation, requires purified primer extension products free from ions and other impurities	Yes
Sauer et al., 2000	GOOD Assay	A primer extension reaction that generates alkylated allele-specific PCR products- alkylation can be more sensitively detected by MS	Sensitive mass spectrometry detection eliminating the rigorous purification steps	Multi-step procedure	Yes

Reference	Method	Concept	Advantages	Disadvantages	PCR step
Chen et al., 2000	Coded Microspheres (Luminex)	Multiplex allelic discrimination reaction performed in solution using 'tagged' primers. 'Tagged' PCR products separated by flow cytometry using complementary 'tagged beads'.	Good multiplex potential	Multiplex PCR and multiplex primer extension reaction are difficult to optimise, and performed in solution separate from solid phase flow cytometry sorting reaction.	Yes
Chen and Kwok, 1999	Template-directed dye-terminator incorporation assay (TDI) using FP or FRET	Donor dye-labelled primer extended using AS-PCR, and acceptor dye labelled ddNTPs. A match/mismatch can be detected using FRET since fluorophores are brought within close enough proximity for the donor to excite the acceptor.	Direct fluorescence detection, cheaper probes, good multiplexing potential, adaptable for FP detection since extension of the donor labelled probe will increase the molecular weight of the fluorophore changing its polarisation properties.	Non-specific products increase noise in the signal.	Yes
Schmaker et al., 1996; Tonisson et al., 2000	Arrayed Primer Extension (APEX), minisequencing	Primer covalently immobilised on microscope slide. Following extension using fluorescent dideoxynucleotides, array analysed by fluorescence scanning	Potential for HTP, uses one detection primer per SNP, ten-fold increase in discrimination potential compared with ASO probe-based array formats.	Requires microarray instruments	Yes

Table 1.3: Common genotyping assays based on allele-specific nucleotide incorporation. The list is not intended to be exhaustive due to the number of similar assays, but rather provide details of a range of assays.

Oligonucleotide ligation	*Based on the discrimination by DNA ligases against mismatches at the ligation site in two adjacently hybridised oligonucleotides*				
Reference	**Method**	**Concept**	**Advantages**	**Disadvantages**	**PCR step**
Lizardi et al., 1998	Rolling circle amplification	Circular DNA ligation products (such as padlock probes) serve as template for a RCA	Signal amplification, yields product many times the size of the original circle. Adapted to SNPgenotyping.	Steric hindrance on solid phase, PCR not required in principle	No
Baner et al., 1998	Padlock probes	Linear oligos with ends complementary to target. When perfectly hybridised, probes are circularised by ligation, acting as templates for RCA.	Localised detection, avoids a PCR step, adapted for SNP genotyping where a mismatch prevents oligo ligation.	Probes difficult to design	No
Whiteley et al., 1989	Ligase chain reaction (LCR)	Two oligonucleotide probes adjacent to a SNP site are ligated in the presence of a matched sequence, forming targets for later cycles	Eliminates a PCR step in principle, yet still accomplishes exponential amplification	Often used in combination with PCR to achieve the required sensitivity (CCR).	No
Bi and Stambrook, 1997	Combined Chain Reaction (CCR)	Consists of denaturation, annealing, extension and ligation, relying on a mismatch at the 5' end of a primer to prevent ligation. Uses thermal stable DNA ligase and DNA polymerase.	Used for SNP detection	Requires post-CCR separation steps on a gel.	Yes
Chen et al., 1998	Dye-labelled oligonucleotide ligation (DOL)	When an allele-specific oligo labelled with an acceptor dye is ligated to an oligo bearing a donor dye, FRET is observed. Cycling conditions are manipulated such that a signal is observed depending on the ligated extended probes.	Performed in a closed tube format with PCR and ligation reactions thermally isolated, made possible by careful primer design, very specific	All 3' ligation probes are labelled with dyes, high reagent cost	Yes

Table 1.4: Common genotyping assays based on allele-specific oligonucleotide ligation. The list is not intended to be exhaustive due to the number of similar assays, but rather provide details of a range of assays.

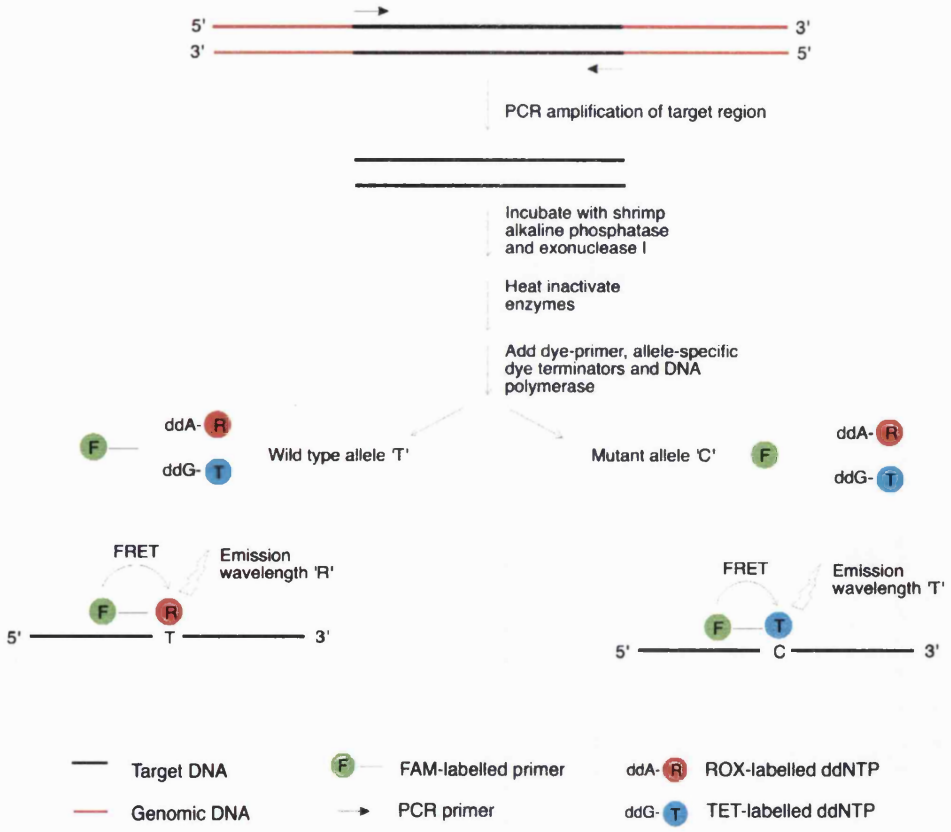


Figure 1.15: The template-directed dye-terminator incorporation (TDI) assay. Target genomic DNA is first amplified by PCR. Excess PCR primers and dNTPs are degraded by *E. coli* exonuclease I and shrimp alkaline phosphatase. Following heat inactivation of these enzymes, the TDI reaction mix is added to the PCR products. An allele-specific primer extension reaction mediated by DNA polymerase occurs in which the donor dye-labelled primer is extended by a single base using a matched acceptor dye-labelled dideoxy terminators. A fluorescent signal characteristic of the reporter labelled ddNTP ensues (Chen and Kwok, 1999).

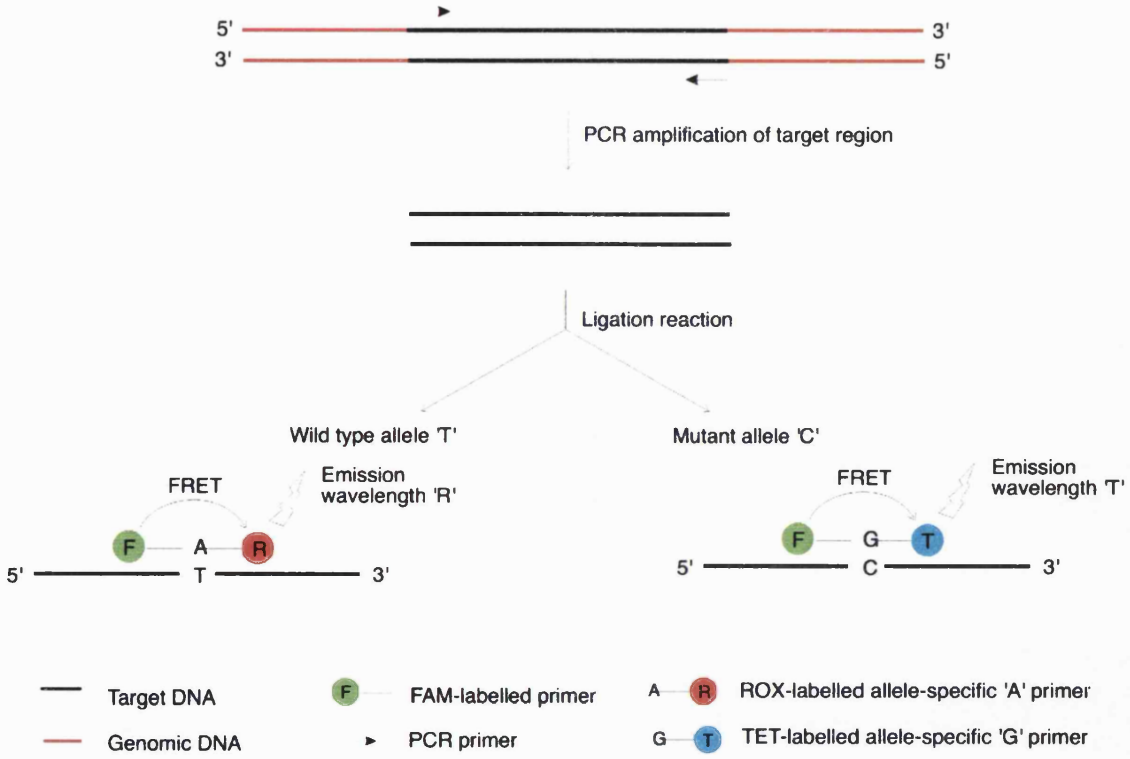


Figure 1.16: Dye-labelled oligonucleotide ligation (DOL) assay with FRET detection. Detection of a T to C SNP is shown in the diagram. A PCR amplification step of the SNP-containing target region achieves the required sensitivity of the assay. This is performed in the presence of sequence-specific dye-labelled probe portions. Upon hybridisation of the allele-specific portion of the probe (two are present, differentially labelled with ROX=R and TET=T), ligation to the FAM (F) labelled portion occurs, bringing the FAM and reporter dye in close enough proximity for FRET to occur. A signal characteristic of the allele-specific portion of the probe ensues allowing allele identification (Chen and Kwok, 1999).

But PCR does have some disadvantages. These include the fact that many clinical laboratories may not have the proper facilities to conduct PCR-based genotyping. Amplicon contamination is a significant and often overlooked limitation for PCR-based techniques. The continual amplification of the same target within a clinic or surgery environment, leads to a significant risk of false positives from air borne amplicon requiring careful disposal of end-point material (Johnson and Johnson, personal communication). In a clinical diagnostic laboratory false positives are highly undesirable, but with the correct controls in place it is possible to recognise contamination should it occur. Physical methods to control contamination can be implemented such as the spatial separation of reagent storage and PCR reaction set-up, sample preparation, PCR reaction mix assembly (i.e. template addition), and PCR-product analysis, each with dedicated equipment (Neumaier *et al.,* 1998). In addition, a licence fee is required to use PCR for diagnostic testing (US Patent No. 4683195 by Cetus Corporation, now owned by Roche).

Consequently, clinical genotyping assays have been developed that can avoid a PCR step, by utilising enzymatic functions other than polymerisation by *Taq* DNA polymerase. A homogenous isothermal (without temperature cycling) reaction format using DNA ligation has been developed for genotyping SNPs, called the ligase chain reaction (LCR, Figure 1.17; Whiteley *et al.,* 1989). Here, two pairs of oligonucleotide probes are used in cyclic ligation reactions, together with a thermostable DNA ligase for exponential amplification of genomic DNA. In practice, DNA ligases are used in combination with PCR to attain the required sensitivity (Combined chain reaction, CCR; Bi and Stambrook, 1997). A summary of other novel isothermal amplification techniques can be found in Table 1.5.

Out of all the nucleic acid amplification techniques described, PCR remains the most widely used and sensitive approach to DNA analysis. Consequently, PCR is the favourable platform on which to base future genotyping assays.

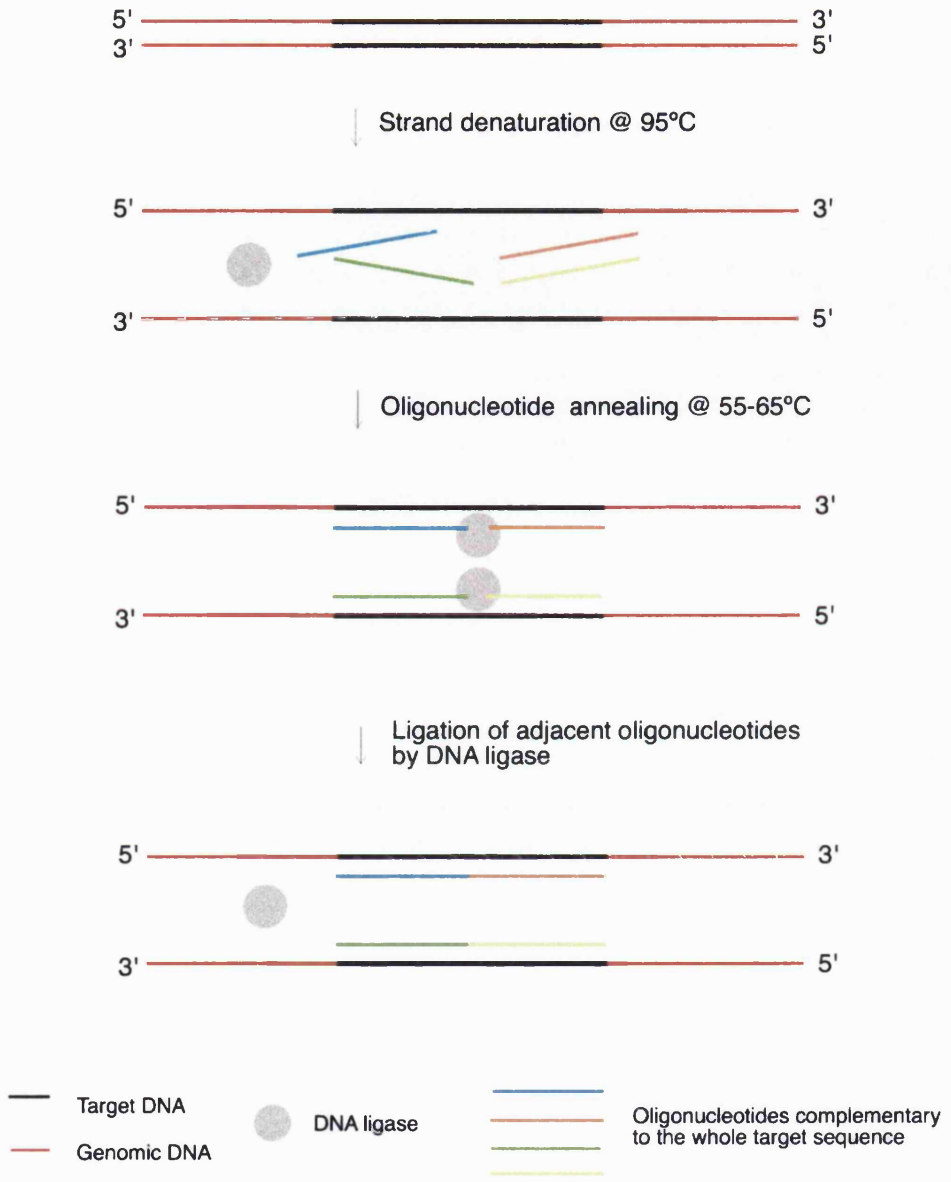


Figure 1.17: The ligase chain reaction (LCR). In LCR, two pairs of probes complementary to the entire target sequence anneal at adjacent positions on the denatured DNA strands. Matched probes will be ligated together by DNA ligase, serving as targets for subsequent cycles leading to exponential amplification (Lee, 1996).

Table 1.5: Common genotyping assays based on isothermal amplification. The list is not intended to be extensive due to the number of similar assays, but rather provide details of a range of assays.

Isothermal assays	*DNA amplification without the need for a PCR step*				
Reference	**Method**	**Concept**	**Advantages**	**Disadvantages**	**PCR step**
Compton, 1991; Guatelli et al., 1990	Nucleic acid sequence-based amplification (NASBA)	Uses reverse Transcriptase, RNAse H and T7 RNA polymerase, in concert with two primers for isothermal amplification at 40°C.	Can give quantitative information about the target gene	Low temperature can result in non-specific primer interactions, not successfully adapted to SNP detection	No
Notomi et al., 2000	Loop-mediated isothermal amplification (LAMP)	Uses DNA polymerase and 4 special primers to amplify DNA by auto-cycling strand displacement DNA synthesis at 60°C.	Can be performed in a laboratory incubator, sensitivity of detection is 6 copies of target gene	Not yet adapted for SNP detection, separate analysis steps	No
Takara Shuzo Co., 2000	Isothermal and Chimeric primer-initiated Amplification of Nucleic Acids (ICAN)	Uses chimera primers (3' end RNA, 5' end DNA), DNA polymerase with strand displacement activity, a ribonuclease that can cut the chimeric primer, and dNTPs, in a template switching reaction.	Has been applied to a chip-based format, takes place at 50-65°C	Not yet adapted for SNP detection	No
Wharam et al., 2001	Signal mediated amplification of RNA technology (SMART) with a three Way Junction (3WJ)	Target dependent signal generated by 3WJ by activation of T7 RNA polymerase promoter. RNA signal measured by enzyme linked oligo-sorbent assay (ELOSA), or real time using molecular beacon.	Assay can be quantitative, requires same reagents regardless of assay and can be used to detect most SNPs.	Complicated probe design, requires improvement in sensitivity	No
Lyamichev et al., 1999; Hessner et al., 2000	Invader Assay (Third Wave Technologies)- *See Figure 1.5E*	A 5' invading oligonucleotide and a partially overlapping 3' signal oligonucleotide form a specific structure when bound to a complementary sequence. A flap endonuclease cleaves the structure, releasing the 5' flap from the probe.	Assay is adaptable for a range of detection methods including mass spectrometry, FRET and FP by labelling the 5' end of the allele-specific oligonucleotide.	Large amount of genomic DNA required to achieve the required sensitivity.	No (in principle)

1.4 Real-time Quantitative PCR methods

Quantitative PCR (qPCR) is a term used to identify any PCR technique that allows determination of the quantity of a specific nucleic acid target in a biological sample. The traditional approach to qPCR is quantitative-competitive PCR (QC-PCR; Becker-Andre and Hahlbrock, 1989), where a competitor amplicon would typically be constructed that contained identical primer binding sites as the target amplicon, yet would be differentiable by size on an agarose gel. The identical binding sites are intended to ensure that the amplification efficiency of both target and competitor amplicons are the same. A known amount of competitor DNA would be spiked into the same reaction, and the assumption made that amplification efficiencies of both amplicons were identical; the ratio of target to competitor should remain the same throughout the PCR test. By determining the ratio of target to competitor at the end of the reaction and knowing the starting amount used to spike the reaction, the starting amount of target can, in theory, be calculated.

There are several drawbacks to this QC-PCR approach. Firstly, quantitative measurements of amplicon are made at the end of a PCR test. Small differences in target concentration, quality and concentration of enzyme, dNTPs, magnesium chloride, primers, cycle length and number, annealing temperature, etc. can greatly affect the amplification efficiency, and subsequently final accumulation of PCR products. This means that even in the best experimental conditions, this and similar end-point approaches can only give a semi-quantitative or qualitative estimation of a nucleic acid template. The dynamic range of a QC-PCR assay is limited to a target-to-competitor ratio of 1:10 or 10:1 meaning that several dilutions must be tested to find a suitable ratio. Further, a suitable competitor requires construction and characterisation for every target to be quantitated (Bustin, 2000).

Quantitative PCR was revolutionised by the advent of real-time PCR (Figure 1.9, pg. 26; Higuchi *et al.*, 1993). In this pioneering research, ethidium bromide was used as a reporter dye but instead of obtaining end-point measurements a kinetic PCR analysis was performed by exciting the PCR tests with UV transillumination,

and taking a video image of the PCR test during each annealing and extension phase using a charge coupled diode (CCD) camera. By monitoring the amplification reactions containing a dilution series of initial starting template and plotting the increase in fluorescence versus cycle number, the quantitative performance of the assay was assessed. It was found that the larger the concentration of starting template copies in the reaction, the earlier the rise in fluorescence above background occurs. This system produces amplification plots that provide a more complete picture of the PCR process than assaying product accumulation after a fixed number of cycles, and allows quantification over a large dynamic range.

PCR instrumentation with the ability to continuously monitor thermal data, combined with sensitive dyes and accurate data analysis software, provide a sensitive platform on which to base homogenous qPCR experiments for a range of clinical applications (Table 1.6). With real-time monitoring, quantification can be performed using data generated during the log-linear (or exponential) phase of PCR (Figure 1.9, Phase 2). During the log-linear phase, amplification efficiency is assumed to be constant, without limiting components, inhibition, or interference from noisy background fluorescence levels. In addition, replicate reactions produce uniform and reproducible results (Rasmussen, 2001).

Monitoring the log-linear phase of a PCR reaction offers significant improvements over methods collating data from plateau phases, which is where most end-point measurements are made. As described earlier, any slight difference in a limiting component can have a drastic effect on the final amount of product, and reproducibility is poor. During the plateau phase, the reaction curve exhibits little or no net increase in PCR product (Figure 1.9, Phase 5). This is thought to be caused by numerous events including depletion of reaction components (dNTPs, enzyme, primers), thermal inactivation of DNA polymerase, inhibition of enzyme activity by increasing pyrophosphate production, reduction in the denaturation efficiency per cycle, destruction of product due to enzyme 5'-3' exonuclease activity, product to product reannealing (Wittwer *et al.,* 1997), the chelation of critical metal ions by the amplicon depriving the enzyme of a cofactor (Wharton and Eisenthal, 1981), or high substrate/amplicon inhibition (Dimitiov and Apostolova, 1995; Kainz, 2000).

Fluorescence readings acquired during the log-linear phase can be used for quantifying nucleic acids in a sample by two basic methods, absolute and relative quantification. For both methods, the generation of reproducible and reliable estimates of nucleic acid concentration relies partly on the accurate measurement of a 'cycle threshold' (C_T) value. The C_T value defines a particular cycle or position of the amplification curve, derived from the point at which the fluorescence signal rises above background noise (Figure 1.9). Identical DNA samples at identical concentrations should produce similar cycle threshold values. The different methods used for defining the C_T using two real-time PCR instruments is summarised in Figure 1.18.

Absolute quantification methods use C_T values to form a standard curve, and target values are usually expressed as an absolute value (copy number or grams). Generally, separate PCR (or reverse transcriptase PCR; RT-PCR, Figure 1.19) tests are run consisting of serial dilutions of known target concentrations (Figure 1.20A), alongside samples of the same target of unknown DNA (or RNA) concentration. The standards are amplified in different reaction vessels, but within the same PCR run (external standards). Higuchi *et al.* (1993) showed that a plot of the log of initial target copy number (or concentration) for a set of standards, versus the cycle threshold is a straight line (Figure 1.20B). Therefore, the cycle threshold value of the unknown can be determined by interpolation from the linear regression of the appropriate external standard curve to estimate the initial DNA (or RNA) concentration.

Application	*Advantages*	*Disadvantages*	*Main considerations*
Clinical, Food and Veterinary Microbiology *e.g. viral load, bacterial load, fungal load*	Similar sensitivity to classical methods but faster, less labour intensive, and higher sample throughput.	Detects nucleic acids from live and dead organisms, classical methods only quantify live organisms. It remains unknown which parameter is most useful in deciding the relevant factor i.e. disease induction, virulence of pathogen.	Infectious agents have high mutation rate, which can influence viral load estimates if primers cannot bind due to mismatches. Therefore, where possible, choose highly conserved regions for assay development.
Clinical Oncology *e.g. minimal residual disease, chromosomal translocations, SNPs*	High technical sensitivity of <5 copies per reaction. Does not entail the primer design issues associated with infectious agents.	Sensitivity is key to minimal residual disease monitoring or transmission of endogenous pathogens during xenotransplantation; therefore maximal input of DNA into the PCR is the limiting step.	Sensitivity and discriminatory abilities are the two main considerations here. The aspects concerned with SNP detection are discussed throughout this Chapter.
Gene Expression and Gene Therapy *e.g. cytokines, receptors etc. or gene transfer estimates, biodistribution of vectors*	Reproducibility of an optimised RT-PCR assay is nearly as good as a real-time PCR assay. Using the comparative quantitative method one can use relative quantification without generating RNA standards.	Need to test and confirm the reproducibility of the housekeeping gene before reliable data can be obtained. Loss of RNA during extraction or storage is common- spiking the sample with a control RNA can test for this problem. Comparative degradation of sample and spiked RNA adds a further issue.	Choice of housekeeping gene and sample preparation is critical. Ensure that DNA does not contaminate the RNA sample otherwise inaccurate gene expression levels will be obtained.

Table 1.6: Selected applications of real-time qPCR including advantages, disadvantages, and main considerations. This list is not intended to be exhaustive (Klein, 2002; Bustin 2000).

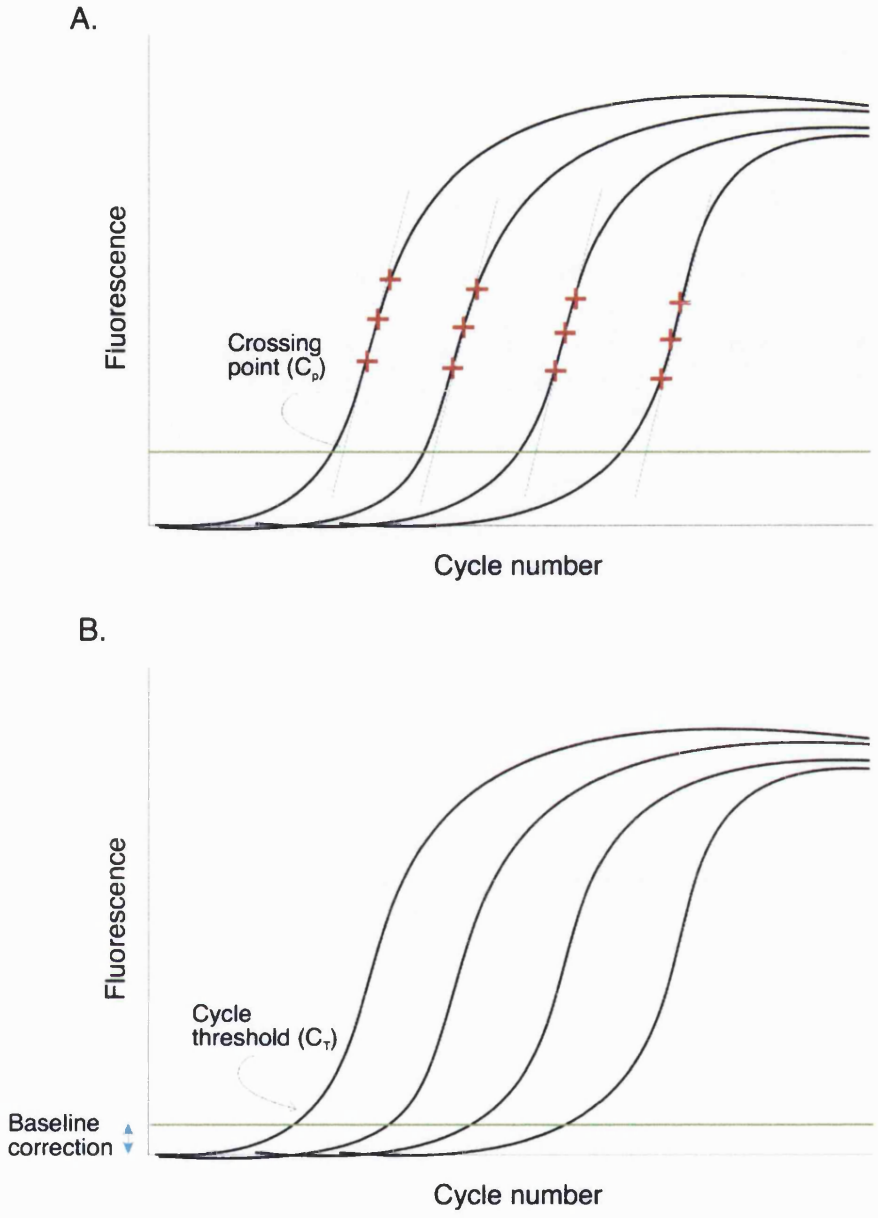


Figure 1.18: [A] LightCycler Fit Points Method of defining the cycle threshold (called the 'crossing point', C_p, by Roche). The user manually sets a baseline (green) above the background noise. Points are selected on the log-linear portion of the curve, and a straight line of $y = mx + c$ is drawn through the points. The crossing point is the fractional cycle number at which an interpolation of the linear curve crossed with the manually set threshold. [B] ABI Prism 7700 Sequence Detection System software automatically performs a baseline correction above the background noise to set a baseline (green line). The fractional cycle number at which a fluorescence reading is detected above this baseline is the cycle threshold (C_T).

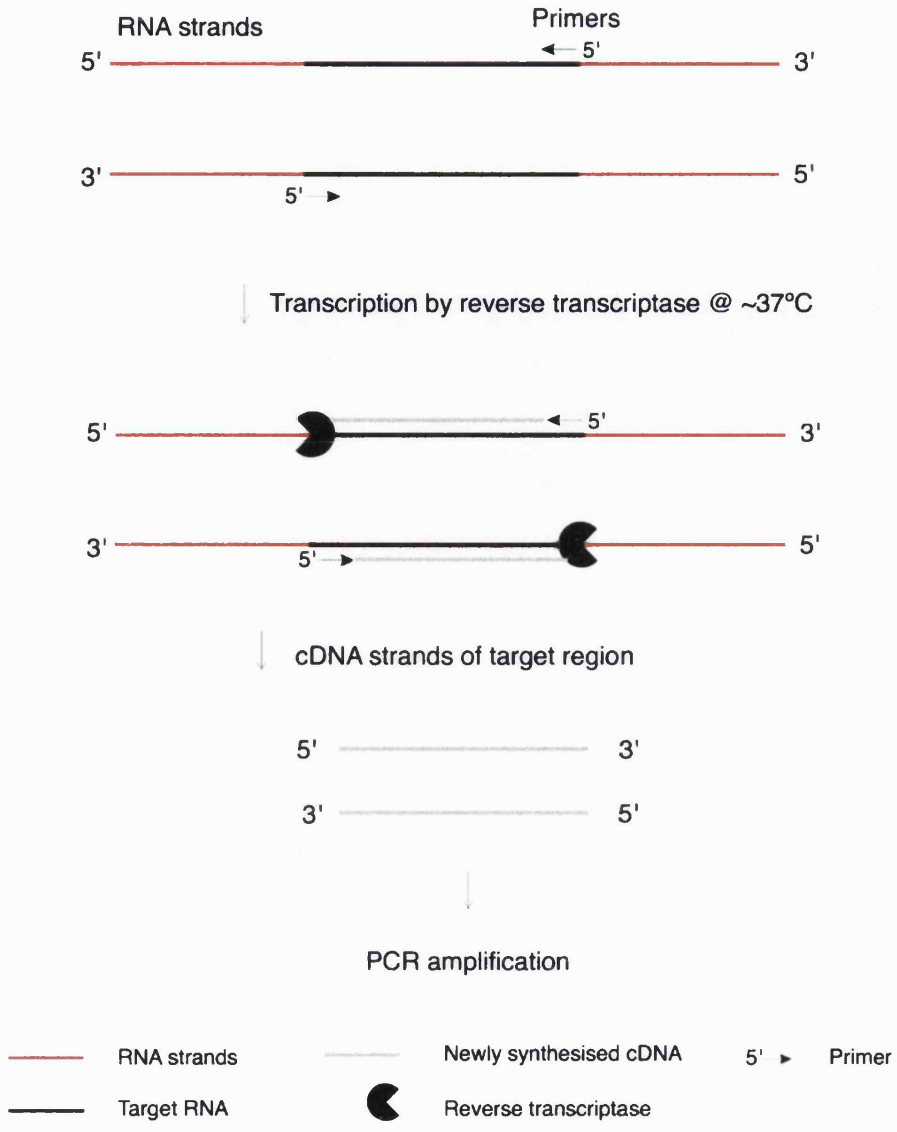


Figure 1.19: Reverse-transcriptase PCR (RT-PCR) for the quantification of RNA (for gene expression studies). An RNA template is transcribed into cDNA by reverse transcriptase. The cDNA is then amplified by conventional PCR.

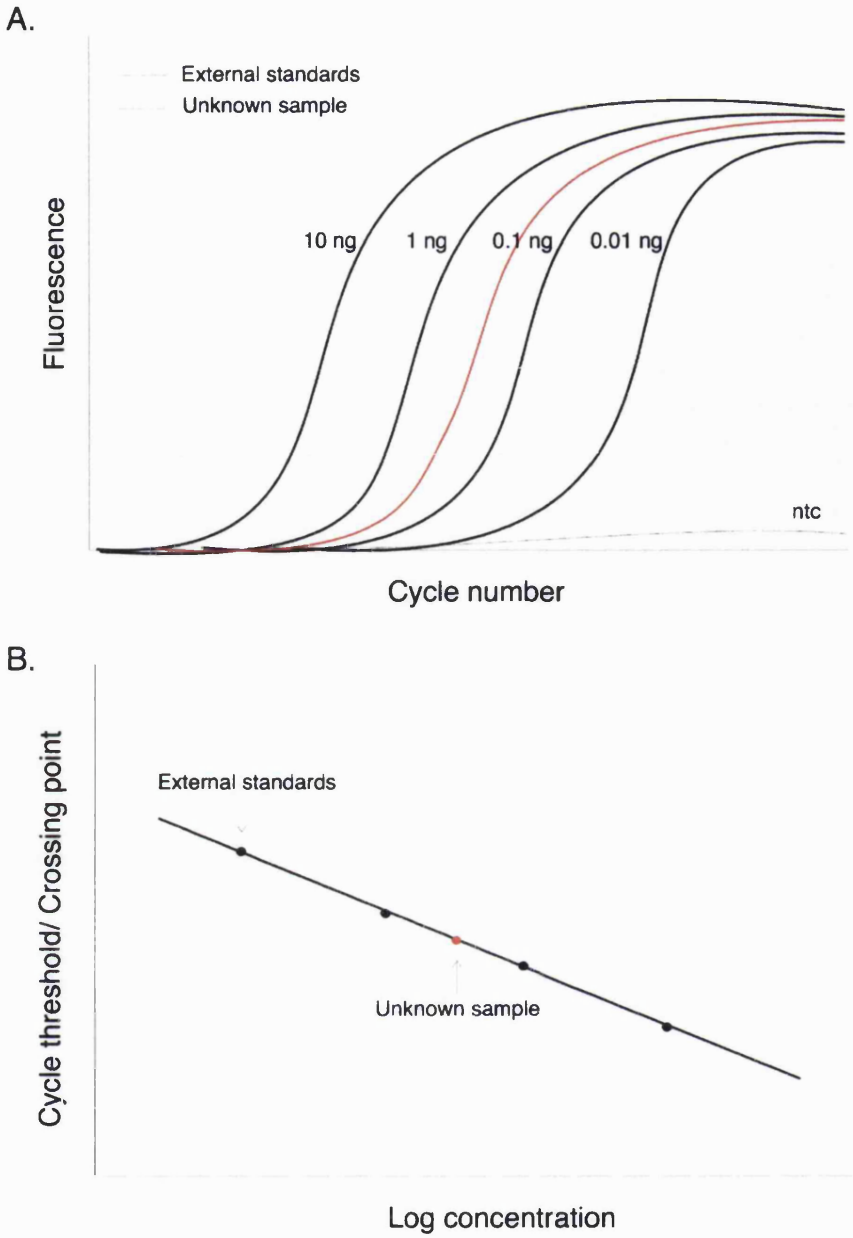


Figure 1.20: Real-time qPCR with external standards using a standard curve and mono-colour detection. [A] Typical amplification profile of product accumulation, monitored using a dsDNA binding dye. Fluorescence readings are acquired at a frequency of once per cycle. For relative quantification, the standards do not have concentration values. Target amplifications are compared with the standard curve from a reference gene (see text) and quantities are given as an n-fold increase in product compared to the reference. [B] A cycle threshold (see Figure 1.15) is defined for each cycle, representing the same point of the PCR curve for each starting DNA concentration. The cycle threshold is plotted vs. the log concentration or log copy number to generate a standard curve from which the concentrations of unknown standards can be interpolated.

There are two key considerations when designing experiments for the absolute quantification of nucleic acid. Firstly, the amplification efficiency of the standards and target DNA must be the same during the run, and between runs (Bustin, 2000). This can be achieved if the targets are identical, or ligated into a plasmid vector meaning the targets have identical binding sites (similar to QC-PCR strategies). However, since the size of the template can also affect PCR efficiency, this could introduce error if genomic DNA samples are being quantified. Secondly, a set of standards containing a defined number of targets is required. Since the 'absolute' numbers are always calculated relative to a standard, the accuracy of the number largely depends on the accuracy of the standard used (Klein, 2002).

Absolute qPCR using external standards can be subdivided into 'mono-colour' detection, or 'dual-colour' detection. The former is accomplished by using a dsDNA specific binding dye (which can be used to monitor all qPCR reactions) or a single real-time PCR probe (introducing sequence-specificity), to monitor the reaction. Using a dsDNA specific binding dye is the least expensive approach, and has a high dynamic range, although there is no control for PCR inhibitors influencing efficiency. When monitoring the test using a dsDNA binding dye, the fluorescence signal may be contributed to by primer dimer amplification as opposed to target. Primer dimer amplification occurs more readily where template concentrations are very low due to the lack of target DNA, and therefore highly optimised reactions are required for sensitive assays (Klein, 2002).

For dual-colour detection, reactions are monitored with an internal control amplification using sequence-specific real-time PCR probes. As with mono-colour detection, the method uses external standards to determine the concentration of the target sample, except the target is co-amplified with a heterologous control amplicon spatially separated from the target. The control is used to measure the PCR performance and test for inhibitors, or inefficient nucleic acid extraction. The control is a desirable attribute to assays used in a clinical setting where the reason for a negative signal cannot be ambiguous (Neumaier, 1998).

Another methodology in qPCR experiments is relative quantification. Relative quantification approaches are used to compare differences in nucleic acid targets among different samples compared to a reference gene, such as monitoring gene expression levels. A reference gene, often called a housekeeping gene or calibrator, is a nucleic acid that is present at a near constant concentration or copy number in all samples being compared. The reference gene may be endogenous, usually a gene that is naturally present in the sample material at constant levels such that the regulation of the target species can be relatively quantified using a second set of primers, or it may be exogenous, an internal reference that is added to the sample at known concentration and is co-amplified with the target nucleic acid using a second set of primers..

Relative quantification may be performed using external references with a standard curve (similar to absolute qPCR methods). In this case, separate tubes are run and the quantity is expressed as an n-fold difference relative to the reference. Instead of an absolute value, for all experimental samples the target quantity is determined from the standard curve and divided by the target quantity of the calibrator (Rasmussen *et al.,* 1998). A typical application of this method would be monitoring drug effects on gene expression using reverse transcriptase PCR (RT-PCR) where an untreated control would be the calibrator and samples isolated from patients using the drug would be compared against this (Klein, 2002).

Alternative methods of relative quantitative PCR are generally designed for the specific software that accompanies particular PCR instruments, using novel algorithms to model the PCR and eliminate the need for standard curve construction. PE Applied Biosystems instruments (Forster City, CA, USA) present a method called the comparative-relative PCR ($\Delta\Delta C_T$). The method is based on the fact that the difference in threshold cycles (ΔC_T) between the gene of interest and the housekeeping gene is proportional to the relative expression level of the gene of interest (Livak and Schmittgen, 2001). Therefore, the target gene quantity can be inferred. The LightCycler™ (Roche Molecular Biochemicals, USA) uses a similar approach, except the relative gene expression ratio is calculated from the real-time

PCR efficiencies and the crossing point (Cp) deviation of an unknown sample versus a calibrator (ΔCp; Pfaffl, 2001).

A pre-requisite for all relative quantification methods is that the PCR efficiency of the target is identical in the reference and sample. In addition, the selection of the reference gene is critical to the success of the assay. For example, in mRNA quantification, a 'housekeeping' gene is used which is a non-regulated gene in the same sample that is constitutively expressed at an identical level to all samples to be analysed. However, even housekeeping genes do not have constant expression levels and as yet, there is no universally accepted method for normalisation using housekeeping genes (Bustin, 2001).

The work described in this thesis builds upon the PCR process to develop assays for the detection of well-characterised SNPs in diagnostic testing. In addition, the kinetics of PCR is explored both qualitatively and quantitatively.

Chapter Two

Materials and Methods

All oligonucleotide sequences (Appendix I) were designed using PrimerCalc© software (Q-Biogene; Carlsbad, CA, USA), and supplied by Sigma Genosys (Gillingham, UK) with a reverse phase cartridge purification step.

2.0 PCR Amplifications

2.0.1 General protocol using an MJ PTC-100 Cycler

For all amplifications performed using a PTC-100 MJ Thermal Cycler (GRI, Essex, UK), Promega (Madison, USA) supplied the reagents. Each PCR was performed in a sterile 0.8 ml polypropylene tube containing 10x Mg-free PCR buffer, oligonucleotides, *Taq* DNA polymerase, magnesium chloride, dNTPs and template DNA. The total reaction volume was made up to 25 µl using PCR-grade water (Sigma Genosys, UK). Component concentrations varied for each experiment, and are stated throughout.

A general thermal cycling program consisted of an initial denaturation step at 95°C for 2 min, followed by 30 amplification cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s. Holding the sample at 72°C for 3 min constituted a final extension step. Protocols deviating from this general program are stated throughput. Samples were visualised on an agarose gel according to Section 2.5.4.

2.0.2 General protocol using a LightCycler™

For all amplifications performed using the LightCycler™ instrument (Roche Molecular Biochemicals, Lewes, UK), Roche Molecular Biochemicals supplied all reagents and consumables. Each PCR was performed in glass capillary, containing DNA Master FastStart SYBR® Green I mixture (10x; *Taq* DNA polymerase, reaction buffer, magnesium chloride, dNTPs, and SYBR Green I dye), oligonucleotides,

additional magnesium chloride, and template DNA. The total reaction volume was made up to 20 µl using PCR-grade water. Component concentrations varied for each experiment, and are stated throughout.

A general thermal cycling program consisted of a 10-minute initial denaturation step (to activate the *Taq* DNA polymerase) at 95°C, followed by 40 amplification cycles of 1 s denaturation at 95°C, 5 s annealing at 60°C, and an extension phase at 72 °C for 5 s. A single fluorescence reading was acquired at the end of the extension phase. All temperature transitions were 20 °C.S^{-1} and fluorescence gains (used to set sensitivity of the LightCycler™) were F1-1 throughout. Protocols deviating from this general program are stated throughout.

2.0.3 Melt curve analysis using a LightCycler™

Following each amplification, a melt curve analysis was performed in the same closed glass capillary with a denaturation step at 95 °C, followed by a cooling step to 5°C above primer annealing temperature (ensuring all fluorescence originates from product-to-product re-annealing) and a slow denaturation phase to 95°C at a rate of 0.1°C.s^{-1}, with continual fluorescence acquisition. Amplicon denaturation was observed as a rapid loss of fluorescence near the calculated melting temperature (T_m). These data were converted into melting peaks by the LightCycler™ Data Analysis Software (Version 3.0; Roche Molecular Biochemicals), by plotting the first negative derivative of fluorescence loss, F, with respect to temperature, T ($-dF/dT$ vs T). The apex of the melt peak is used to represent the specific T_m of the product.

2.1 Plasmid DNA Library Construction

A small library of four plasmids was prepared to model a "susceptibility" gene. Each plasmid is identical apart from the presence of a single nucleotide polymorphism (G, C, A or T) at identical base positions. Promega supplied all reagents unless otherwise stated.

2.1.1 Amplification of target sequences containing polymorphism

PCR tests were performed on the LightCycler™ according to Section 2.0.2, except an extension time of 10 s was used. Amplification was directed by primers ACT BF and ACT BR generating a 374 bp insert from human genomic DNA. Primers were included at a concentration of 0.6 µM each, at a final magnesium chloride concentration of 3 mM. Human genomic DNA template was included at 15 ng total. A melt curve analysis (Section 2.0.3) generated a single peak at 88 °C, confirming specificity of the product.

PCR amplicons were purified using Wizard® PCR Preps DNA Purification System (Section 2.3.1), and subsequently used as template DNA in PCR reactions directed by primers ACT BF and ACT BRX (where X is G, C, A or T), according to test conditions described above. Primer ACT BRX was designed with a single polymorphism at a defined position in a 5' tail, to incorporate a SNP at the same position in the four resulting amplicons. Amplicons were purified, and stored at –20°C until use.

2.1.2 Ligation of PCR fragments into plasmid vector

Ligations were carried out using the pGEM®-T Easy Vector System II. The linearised vectors have a 3' terminal thymidine residue at both ends complementary to the poly 'A' tail added to PCR amplicons synthesised by *Taq* DNA polymerase. 5 µl of 2X Rapid Ligation Buffer was added to 1 µl of pGEM®-T Easy Vector (50 ng), 1 µl of T4 DNA Ligase (3 Weiss units.μl^{-1}), and 7 µl of purified PCR product. A positive control and background control were prepared by replacing PCR fragments with Control Insert DNA or PCR-grade water, respectively. All reactions were mixed by

pipetting, and incubated for 1 hour at room temperature. Figure 2.1 displays the plasmid map. The position and sequence of the cloned insert is described in detail in Appendix II. The position of the polymorphic sequence was verified by sequencing using BigDye Terminator Cycle Sequencing Ready Reaction Kits on an ABI Prism 377 DNA Sequencer (Applied Biosystems; Foster City, CA, USA).

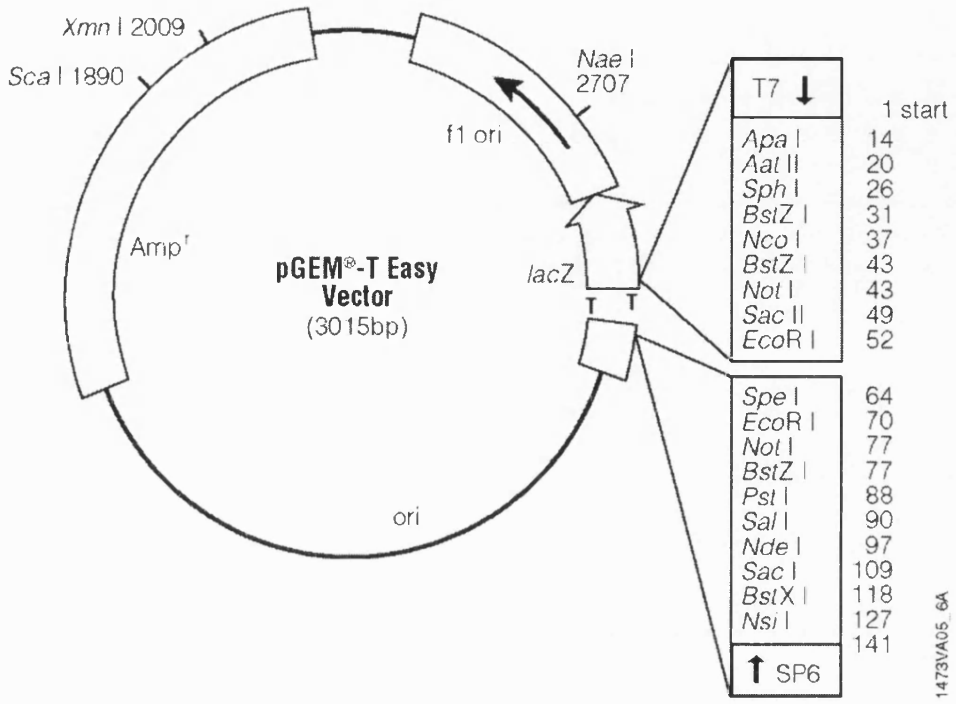


Figure 2.1: Plasmid map of the pGEM-T® Easy Vector Plasmid System II. The β-actin DNA insert is in the multiple cloning site at position 52. Actual sequence is shown in Appendix II. Diagram reproduced with permission from http://www.promega.com/vectors/cloning_vectors.htm.

2.2 Preparation and Transformation of Competent *Eschericia coli* Cells

All chemicals and reagents were supplied by Sigma Aldrich (Dorset, UK) and were of analytical grade unless otherwise stated. Becton Dickinson and Company (New Jersey, USA) supplied all media chemicals unless otherwise stated.

2.2.1 Preparation of fresh or frozen competent E. coli cells

Competent *E. coli* cells were routinely prepared from a frozen stock (stored at –70°C in 15% sterile glycerol) by streaking the *E. coli* cells onto the surface of a Luria-Bertani (LB media: 10 $g.l^{-1}$ Bacto®-tryptone, 5 $g.l^{-1}$ Bacto®- yeast extract, 5 $g.l^{-1}$ NaCl) agar plate (15 $g.l^{-1}$ agar; Oxoid, Hampshire, UK) using a sterile platinum wire. The plates were incubated for 12-16 hours at 37°C.

5 ml LB media (10 $g.l^{-1}$ Bacto®-tryptone, 5 $g.l^{-1}$ Bacto®- yeast extract, 5 $g.l^{-1}$ NaCl) was inoculated with a single well-isolated colony of 1-2 mm diameter, and grown for 3 hours at 37°C. Cells were transferred aseptically into sterile, ice-cold 50 ml polypropylene tubes and cooled to 0°C by storing on ice for 10 min. Cells were recovered by centrifugation at 4000 rpm for 10 min at 4°C. Media was decanted from the cell pellets, and tubes were inverted for 1 min allowing the last traces of media to drain away. Pellets were resuspended in 20 ml (per tube) of ice-cold transformation buffer (TFB; 10 mM MES [pH 6.3], 45 mM $MnCl_2.4H_2O$, 10 mM $CaCl_2.2H_2O$, 100 mM KCl, 3 mM Hexamminecobalt chloride). When frozen competent *E. coli* for storage at –70°C were prepared, FSB (10 mM Potassium acetate [pH 7.5], 45 mM $MnCl_2.4H_2O$, 10 mM $CaCl_2.2H_2O$, 100 mM KCl, 3 mM Hexamminecobalt chloride, 10% glycerol) was used in place of TFB. Resuspended cells were stored on ice for 10 min, and recovered by centrifugation at 4000 rpm for 10 min at 4°C. The supernatant was decanted and tubes were inverted for 1 min to remove the last traces of buffer. The pellet was resuspended in 4 ml (per tube) of ice-cold TFB or FSB.

To prepare fresh competent cells, 140 µl of DnD solution (1.53 g dithiothreitol, 9 ml DMSO, 100 µl 1 M potassium acetate [pH 7.5], H_20 to 10 ml) was added to each suspension and mixed gently by swirling; the suspension was stored on ice for a further 15 min. Aliquots of the suspensions were dispensed into chilled, sterile polypropylene tubes, and stored on ice before use.

To prepare frozen stocks of competent cells, 140 µl of DMSO was added per 4 ml of resuspended cells. Tubes were mixed gently by swirling and stored on ice for 15 min. An additional 140 µl of DMSO was added to each suspension and mixed gently before returning the suspensions to an ice-bath. 400 µl aliquots were dispensed into chilled, sterile microfuge tubes and immediately snap-froze in liquid nitrogen prior to storage at –70°C until required.

2.2.2 Transformation of fresh or frozen competent E. coli cells

Plasmid DNA was added to 10 µl of thawed (from frozen stock) or fresh competent *E. coli* cells (prepared in Section 2.2.1). Tubes were gently swirled several times to mix their contents and stored on ice for 30 min. The volume of DNA solution did not exceed 5% of the volume of competent cells. Two controls were incorporated including competent bacteria that receive a known amount of the standard preparation of supercoiled plasmid DNA and competent bacteria that receive no plasmid DNA at all.

Tubes were heat-shocked to 42°C in a water bath for exactly 2 min, and rapidly transferred to an ice bath to chill for 2 min. 1 ml of LB media was added to each tube before being placed in a shaking incubator set at 37°C for 45 min, allowing bacteria to recover and express the antibiotic resistance marker encoded by the plasmid (ampicillin resistance). Cells were concentrated by centrifugation (4000 rpm for 10 min at room temperature) and the supernatant was discarded. The pellet was resuspended by gentle pipetting ready for colony selection by insertional inactivation (Section 2.2.3).

2.2.3 *Testing bacteria for insertional inactivation*

40 µl of a stock solution of X-gal (20 $mg.ml^{-1}$ in dimethylformamide) and 4 µl of a solution of isopropylthio-β-D-galactodise (IPTG) (200 $mg.ml^{-1}$) were added to an LB agar plate containing 50 $\mu g.ml^{-1}$ of ampicillin. A sterile glass spreader was used to spread the solutions over the agar surface and plates were incubated at 37°C until the fluid had disappeared. The supplemented LB plate was inoculated with 200 µl of the suspension to be tested, and left until the inoculum had been absorbed. The plate was incubated in an inverted position for 12-16 hours at 37°C.

Plates were stored at 4°C for several hours until the blue colour had fully developed. All colonies were transformants since they contained the active antibiotic resistance gene. The ability of the β-galactosidase to form blue colonies in the presence of IPTG shows that the *lacZ* gene was not interrupted by successful ligation of the PCR insert. Positive white colonies demonstrated that the insert was present. Single colonies were picked and spread onto fresh LB X.I.A agar plates, and grown overnight at 37°C to confirm that colonies are actually white and not slow growing blue colonies (Sambrook *et al.,* 1989).

2.3 DNA Purification

Buffer contents from DNA Purification Kits are disclosed, unless information is proprietary to the Supplier.

2.3.1 Wizard® PCR Preps for DNA purification

The Wizard® PCR Preps DNA Purification System (Promega) was used to isolate PCR amplicons from non-specific artefacts including primer dimer. 50 µl of completed PCR test was added to a clean 1.5 ml microcentrifuge tube containing 100 µl of Direct Purification Buffer (50 mM KCl, 10 mM Tris-HCl [pH 8.8], 1.5 mM $MgCl_2$, 0.1% Triton® X-100). The mix was vortexed briefly. Guanidinium chloride contain resin (1 ml) was added and vortexed briefly three times over a 1-minute period. For each PCR product, one Wizard® Minicolumn was prepared with the tip inserted into a vacuum manifold. The resin/DNA mix was pipetted into the syringe barrel, and a vacuum was applied to draw the resin/DNA mix through the Minicolumn. The column was washed by drawing 2 ml of 80% isopropanol through the syringe barrel under vacuum, and dried by drawing a vacuum for a further 30 seconds after the solution had been pulled through. The syringe barrel was transferred to a 1.5 ml microcentrifuge tube and residual isopropanol was removed by centrifugation at 4000 rpm for 2 min. The Minicolumn was transferred to a new centrifuge tube. 50 µl of PCR-grade water was applied to the Minicolumn and left to stand for 1 min before centrifugation at 4000 rpm for 20 seconds to elute the DNA fragment. The purified DNA was stored at 4°C until use.

2.3.2 QIAprep® Spin Miniprep Kit protocol

The QIAprep® Spin Miniprep Kit (Qiagen Sciences, Maryland, USA) was used to purify the plasmid DNA. Overnight cultures of transformed *E. coli* were prepared by inoculating 5 ml of LB media containing 50 $\mu g.ml^{-1}$ of ampicillin with a single positive colony (prepared as described in Section 2.2.3). Solutions were incubated for 12-16 hours at 37°C. Bacterial cells were concentrated by centrifugation at 10000 rpm for 2 min. The media was discarded and the bacterial pellet was resuspended in 250µl Resuspension Buffer (Buffer P1; 50 mM Tris.Cl, pH 8.0, 10 mM EDTA, 100 $\mu g.ml^{-1}$

Rnase A) to weaken the bacterial cell wall. The suspension was transferred to a 1.5 ml microfuge tube and 250 µl Lysis Buffer (Buffer P2; 200 mM NaOH, 1% SDS w/v) was added. The solution was mixed gently by inversion. The alkaline conditions completely lysed the phospholipids and protein components of the cell membrane.

350 µl Buffer N3 (containing 25-50% guanidinium chloride and 10-25% acetic acid) was added and tubes inverted 4-6 times, until the solution became cloudy. This neutralised the whole lysate causing chromosomal DNA to denature forming an insoluble network, whilst protein-SDS complexes and high molecular weight RNA precipitated. Samples were centrifuged at 4000 rpm for 10 min until the supernatant containing plasmid DNA was clear and a compact white pellet of the contaminating macromolecule was formed. Supernatant was removed by gentle aspiration and applied to the QIAprep® spin column. Tubes were centrifuged for 1 min and the flow-through was discarded. Buffer PB (0.5 ml containing 25-50% guanidinium chloride and 25-50% propan-2-0l) was added to the column, followed by centrifugation for 1 min to remove trace nuclease activity.

The QIAprep® spin column was washed by adding 0.75 ml Buffer PE (containing ethanol) and centrifuging for 1 min. Flow-through was discarded, and an additional centrifugation step (1 min) removed residual wash buffer. Finally, the QIAprep® column was placed in a clean 1.5 ml microfuge tube and DNA was eluted by adding 50 µl PCR-grade water to the centre of each column. Columns were left to stand for 1 min prior to centrifugation at 4000 rpm for 1 min.

2.4 Plasmid DNA Quantitation

Plasmid DNA was quantified following purification. It is essential that concentrations are identical in order to validly compare results obtained using each plasmid DNA template. Since spectrophotometric analysis (Section 2.4.2) often overestimates DNA yield and can be slightly variable, concentrations were refined by real-time quantitative PCR methods (Section 2.4.3).

2.4.1 Agarose Gel Electrophoresis

10 µl of Lambda DNA (Promega) concentration standards (5 to 50 $ng.\mu l^{-1}$) with 2 µl bromophenol blue gel loading buffer were separated electrophoretically alongside unknown DNA samples on a 1% agarose gel (Molecular Biology Grade Agarose; Q-Biogene) containing 0.5 $\mu g.ml^{-1}$ ethidium bromide (Sigma Aldrich). Electrophoresis was performed in 1x TBE (5x TBE stock solution; 0.445 M Tris-borate, 10 mM EDTA, [pH 8.3]) at 30 mA constant current. Products were visualised by UV transillumination (305nm) using a High Performance Ultraviolet Transilluminatior (Ultra-Violet Products Limited; Cambridge, UK) and images captured. Maximal band optical density (OD_{max}) was measured using LabWorks™ Gel Analysis Software (Version 3.0; Ultra-Violet Products Limited) and plotted as a standard curve, allowing estimation of unknown DNA concentrations.

2.4.2 Spectrophotometric Analysis

DNA was quantified by spectrophotometric analysis using a Cary-100 Bio UV-Visible Spectrophotometer (Varian; Surrey, UK). 150 µl of PCR-grade water was added to a microcuvette with a path length of 0.5 cm and baseline absorbance readings were obtained between 260 nm and 280 nm. Purified plasmid DNA (Section 2.3.2) was diluted into PCR-grade water up to 150 µl, and added to a clean quartz microcuvette. Absorbance readings from 260 nm to 280 nm were taken using the baseline correction. The reading at 260 nm was recorded and used to calculate the concentration of nucleic acid in the sample since 1 OD corresponds to ~ 50 $\mu g.ml^{-1}$ (average extinction coefficient) for double stranded DNA. Absorbance

readings below 0.05 at 260 nm were considered insignificant and the OD_{260}/OD_{280} value assessed DNA purity, which was ≥ 1.7 in each case (Sambrook *et al.*, 1989).

2.4.3 Quantitative PCR (qPCR) Methods

Plasmid concentrations estimated by spectrophotometric methods (Section 2.4.2), were refined using real-time qPCR methods. Primers ICF and ICR were used to direct the amplification of a 223 bp fragment from plasmid DNA, spatially separated from the DNA insert and variant nucleotide. Each PCR was performed on the LightCycler™ according to Section 2.0.2, containing 0.5 μM each primer, and a final concentration of 3 mM $MgCl_2$. External standards consisted of 2 μl of plasmid preparation (with variant nucleotide G) at an initial estimated concentration of 10 $ng.\mu l^{-1}$, and 10x dilutions (1:10, 1:100, 1:1000, and 1:10000). These were amplified alongside tests containing 2μl of plasmid preparations C, A, and T. Following PCR, a melt curve analysis was performed (Section 2.0.3), to confirm product specificity. The Fit Points Method (LightCycler™ Data Analysis Software; Version 3.0) was used to estimate initial concentrations of plasmids relative to the external standard. All plasmids were then diluted to the same concentration as the standard.

2.5 Human Genomic DNA Isolation and Restriction

Buffer contents from DNA Isolation Kits are disclosed, unless information is proprietary to the Supplier.

2.5.1 Isolation from Buccal Cell Swab

Human genomic DNA was isolated from buccal cell swab using the Puregene® DNS Isolation Kit (Gentra Systems, Minneapolis, USA). Buccal cell samples were collected by scraping the inside of mouth with a sterile nylon cytology brush. The collection brush was dipped up and down 10 times in a 1.5 ml microfuge tube containing 300 µl Cell Lysis Solution, and incubated at 65°C for 60 min. 1.5 µl of RNAse A solution was added to the cell lysate followed by mixing by inversion 25 times. The sample was incubated at 37°C for 15-60 min. The sample was cooled to room temperature and 100 µl of Protein Precipitation solution was added to the cell lysate. The tube was placed in an ice bath for 5 min before centrifugation at 10 000 rpm for 3 min. The supernatant containing the DNA was poured into a clean 1.5 ml microfuge tube containing 300 µl 100% isopropanol and 0.5 µl Glycogen solution (20 $mg.ml^{-1}$). The sample was mixed by inversion 50 times and incubated at room temperature for 5 min. The sample was centrifuged at 10000 rpm for 5 min and the DNA was visible as a small white pellet. The supernatant was discarded, and the tube drained on clean absorbent paper. 300 µl of 70% ethanol was added to wash the DNA followed by centrifugation for 1 min. The ethanol was carefully poured away and the tube left to air-dry for 15 min. 20 µl of DNA Hydration Solution (10 mM Tris, 7 mM EDTA [pH 7.0-8.0]) was added and incubated for 1 hour at 65°C to disperse the DNA. DNA was stored at 4°C until use.

2.5.2 Requesting DNA samples from external laboratories

DNA samples isolated from patients with sickle cell anaemia or hereditary haemochromatosis were obtained from external laboratories. A letter was sent to potential suppliers (Figure 2.2), for approval by the Ethical Committee of the appropriate institution. Samples were assigned a number and these were the only

details given to ensure patient confidentiality. The sample genotypes according to the source laboratory were only supplied following in-house tests.

2.5.3 DNA restriction, Agarose Gel Electrophoresis

1 µg DNA was restricted with 20 units restriction enzyme in One-Phor-All *Plus* buffer (Amersham Pharmacia, Little Chalfont, UK) made up to a final volume of 30 µl with PCR-grade water. Samples were incubated for 1 hour at the appropriate temperature followed by an enzyme deactivation step for 15 min. Digestion products (10 µl) were added to bromophenol blue gel loading buffer (2 µl) and separated electrophoretically on agarose gels (0.7% to 2.5%) containing 0.5 $\mu g.ml^{-1}$ ethidium bromide. Electrophoresis was performed in 1x TBE at 30 mA constant current. Products were visualised by UV transillumination on a High Performance Ultraviolet Transilluminatior (305nm). Gels were photographed and analysed by LabWorks™ Gel Analysis Software (Version 3.0).

Dear Sir/Madam

Re: DNA samples for SNP detection

Further to our telephone conversation on the 7th February, I am writing with a request for some DNA samples to be used for research purposes only.

I am a final year PhD student at the University of Bath, sponsored by Molecular sensing plc, working on the development of novel assays for single nucleotide polymorphism (SNPs) detection. My work so far has culminated in the development of a PCR-based assay enabling the rapid identification of sickle cell anaemia from human genomic DNA samples (provided by Dr. John Old at the National Haemoglobinopathy Reference Service, Oxford, UK).

However, I have been advised that for a more thorough evaluation of this test, I need to demonstrate its applicability to other polymorphisms/deletions, such as those contributing to diseases such as cystic fibrosis, hereditary haemochromatosis, familial hypercholesterolaemia, and other genetic disorders. Therefore, I am writing to ask if you could kindly spare a small amount of DNA from approximately 10-15 patients with a disease from the list above (or any other clinically relevant SNP), which I could use for the purpose of assay validation to help in my PhD research.

On the advice of the Cytogenetics laboratory at Southmead, I have contacted Simon Ramsden at St. Marys Hospital, Manchester, to check the regulations for sending DNA samples between laboratories prior to sending this request. He advised that currently this is not a problem, but within the next few months a restriction will be put in place. He could not see any problems for DNA to be transferred at this time.

I have enclosed details of my two PhD supervisors below such that they can be contacted if appropriate. Also, I would be more than happy to visit your laboratory to discuss my request further.

Please do not hesitate to contact my supervisors or me if you have any queries and many thanks in anticipation of your help.

Yours sincerely

Christy Waterfall

Figure 2.2: Example letter requesting DNA samples for SNP detection.

Chapter Three

Optimisation of Allele-Specific PCR (AS-PCR) for SNP Genotyping

3.0 Introduction

Prior to polymerase chain reaction development (Chapter One; Mullis and Faloona, 1987, Saiki *et al.,* 1985), diagnostic tests for the detection of single base polymorphisms were restricted, primarily due to a lack of sensitivity and labour-intensive analyses (Chang and Kan, 1982). Typically, each test required approximately 250 ng of genomic as DNA starting material to acquire informative results, which represents the upper level of most extraction techniques. PCR allows the exponential amplification of specific regions of DNA copies from a single genomic copy, delivering the desired sensitivity and specificity for a range of genotyping tests including traditional RFLP-based assays (Figure 1.5F, pg. 18) to contemporary sequence-specific detection using oligonucleotide probes (Figure 1.5A). Consequently, the majority of all genotyping tests developed to date involve a PCR step, either front-end, back-end or integral to the test (Tables 1.1-1.5 inclusive, pgs. 30, 44, 46, 47, 54 and 58).

In 1989, PCR itself was modified for SNP detection in a technique called allele-specific PCR (AS-PCR), also referred to as the amplification refractory mutation system (ARMS; Newton *et al.,* 1989), allele-specific amplification (ASA; Nichols *et al.,* 1989), or preferential amplification of specific alleles (PASA; Sarkar *et al.,* 1990). A schematic of conventional AS-PCR is shown in Figure 3.1. The design of conventional AS-PCR tests is simple. Typically two separate PCR tests are required to genotype a bi-allelic system. One test contains an allele-specific primer with complementarities to the wild type DNA sequence and the other contains an allele-specific primer with 3' terminal complementarities to the mutant DNA sequence. Given that *Taq* DNA polymerase initiates synthesis from the 3' end of a primer and lacks a 3'→5' proofreading activity, a mismatched primer is refractory to extension under optimised conditions, resulting in poor amplification (Newton *et al.,* 1989).

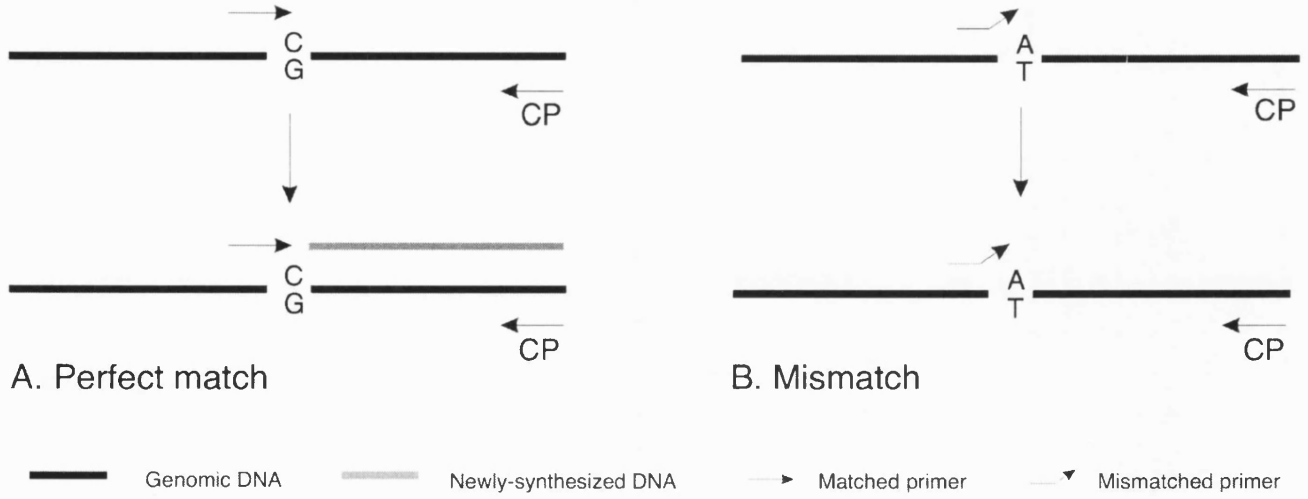


Figure 3.1: Schematic representation of conventional allele-specific PCR for the detection of a C to A polymorphism. [A] The allele-specific primer (with 3' terminal G) and common primer (CP) will amplify from genomic DNA samples homozygous for the wild type 'C' allele. A newly synthesised DNA strand of a pre-defined size will be produced. [B] Genomic DNA samples homozygous for the mutant allele 'A', genomic DNA will not be amplified efficiently due to a mismatch at the 3' end of the wild type allele-specific primer. The genotype can be inferred by performing two separate PCR tests incorporating the same template DNA containing each allele-specific primer complementary to the wild type or mutant allele. Heterozygotes will synthesise the correct amplicon in both reactions since both alleles are present.

The inclusion of a common reverse primer, designed to lie at a pre-defined distance from the SNP site, allows the user to identify the allele-specific products, typically by size separation of DNA fragments methods and subsequent staining with a DNA binding dye such as ethidium bromide (Figure 1.6A, pg. 20; Markovits *et al.,* 1979). The genotype can be inferred from amplicon profiles of the two separate PCR tests. The wild type primer is refractory to extension in the presence of mutant DNA, yet forms a product of the correct size in the presence of wild type DNA, and the mutant primer is refractory to extension in the presence wild type DNA, yet forms a product of the correct size with mutant DNA. In the case of heterozygous inheritance of a particular SNP, products of the pre-defined size are amplified in both tests. An internal control set of primers (non-allele specific and spatially separated from the region of interest) is typically included in both PCR tests to ensure that a negative result is solely caused by a mismatched primer, and not by a PCR inhibitor or insufficient starting DNA for example (Bottema and Sommer, 1993).

This chapter describes the development of a single-tube AS-PCR test. The test, called bi-directional PCR, combines two allele-specific primer sets in the same tube, arranged in a bi-directional orientation around the SNP site, with complementarities to either wild type or mutant alleles. Outer primers are positioned at different distances from the polymorphic site to allow genotype identification by size discrimination on an agarose gel. Other single tube adaptations of AS-PCR have recently been described (Table 3.1). They differ from the bi-directional PCR assay described here by employing varied design features to achieve specificity and detect the amplicons.

AS-PCR tests are simple to design from known sequence information, satisfying a criterion for the "perfect assay" for genotyping (Figure 1.4, pg. 16). However, it can be challenging to establish conditions that are truly discriminatory for all SNP combinations. This is particularly significant where end-point measurements are used for product analysis since a single misextension event will generate a matched template for subsequent cycles. The complementary template can be amplified to plateau phases in just a few cycles meaning that match- and mismatch- primed amplicon profiles are indistinguishable, causing failure of the genotyping test.

Table 3.1: A comparison of the single-tube allele-specific PCR assays published to date that do not utilise fluorogenic probes. These include competitive oligonucleotide priming (Gibbs *et al.,* 1999), tetra-primer PCR (Ye *et al.,* 1992), Bi-PASA (Lui *et al.,* 1997), Tm-shift genotyping (Germer and Higuchi, 1999), SAS-PCR (Sasvari-Szekely *et al.,* 2000), Tetra-primer ARMS-PCR (Ye *et al.,* 2001), and PCR-CTPP (Hamajima, 2001). These are compared to bi-directional PCR (Waterfall and Cobb, 2001) shown in **bold** type. AGE represents agarose gel electrophoresis.

Method	*COP*	*Tetra-primer PCR*	*Bi-PASA*	*T_m-shift genotyping*	*SAS-PCR*	*Tetra-primer ARMS PCR*	*PCR-CTTP*	***Bi-directional PCR***
Total number of primers	3	4	4	3	4	4	4	**4**
Inner primers	2	2	2	2	2	2	2	**2**
Allele-specific mismatch	Internal	Internal	3'	3'	3'	3'	3'	**3'**
Additional mismatch	No	No	No	No	Yes (-2 from 3' terminus)	Yes (-2 from 3' terminus)	No	**No**
Length, bases	~20	~15	~20	~24	~22	~28	~21	**~20**
GC Tail	No	No	Yes (10 bases)	Yes (26 bases)	No	No	No	**No**
Inner/outer primer ratio	1:1	35:1	1:1	1:1	1:1	10:1	1:1	**Variable**
Target region as control amplicon	N/A	No	Yes	N/A	Yes	Yes	Yes	**No**
Annealing temperature	Constant	Higher in earlier cycles	Constant	Constant	Constant	Constant	Constant	**Constant**
Detection method	Autoradiography	Size by AGE	Size by AGE	Melt curve analysis	Size by AGE	Size by AGE	Size by AGE	**Size by AGE**
Special feature	Using primer competition and differentially labelled allele-specific oligo for product identification	Melting temperatures of inner and outer primers vary by 10°C to allow high stringency cycling	Uses a GC tail to control stringency of test	Uses a GC tail on one of the allele-specific primers to increase the Tm of associated product	Uses additional mismatches to provide specificity	Has available primer design software to improve automation	Validated using many different SNP models	**Compatible with multifactorial optimisation strategy for improved discrimination**

The method used for optimisation and improved discrimination should be quick and easy to implement to reduce the overall costs associated with the genotyping test. A common approach to improving the likelihood of discrimination is the incorporation of additional mismatches at the antepenultimate 3' position of the allele-specific primer to further weaken hydrogen bonding with the template. Generally, a mismatch at this position will not be sufficiently disruptive to prevent primer extension if the 3' terminal base is matched, but in the presence of mutant template where both positions are mismatched, severe destabilisation of the primer-template duplex can occur, preventing extension (Sommer *et al.,* 1992; Hézard *et al.,* 1997). Table 3.2 summarises data from published reports employing this approach to detect all SNP combinations. Frequently mismatch extension was prevented, however only one group out of the seven report conditions in which all SNP combinations were effectively discriminated (Bottema *et al.,* 1993).

Indeed, the ability to define discriminatory conditions is highly dependent upon a number of physical parameters including PCR annealing temperature, primer length, instrumentation (Table 3.3), PCR enhancers (Table 3.4), and the reagents and enzyme system used. A significant improvement in the quality of primer and enzyme systems has occurred over the past decade, primarily due to new purification and synthesis methods (Linz *et al.,* 1990). In particular, the advent of *"hot start"* enzyme systems has dramatically improved reaction specificity. Hot Start systems are designed to eliminate the spurious amplification caused by *Taq* polymerase activity occurring at ambient temperatures during reaction set up. Primers can non-specifically anneal to the template below the temperature for specific annealing and cause failure of genotyping tests. Numerous methods have been described to prevent this non-specific extension, mainly by debilitating the enzyme at low temperatures (<40°C; Table 3.5).

Table 3.2: Summary of Prior Studies reporting effects of Primer 3' nucleotide mismatches on the efficiency of amplification by *Taq* DNA polymerase in PCR. For purposes of comparison, data has been interpreted in a roughly quantitative way by a scale of (+++++) for equivalent to a matched amplification, to (-) for no amplification. For Newton *et al.,* 1989, results were given as 'refractory' represented by (-), or not refractory, represented as the equivalent of a match (+++++). NC denotes No Comment on this particular combination. Allele-specific PCR was used, followed by gel electrophoresis with subsequent ethidium bromide staining. Band intensities were qualitatively compared. For Kwok *et al.,* 1990, HIV I model system was used in allele-specific PCR tests, quantified by ethidium bromide staining following gel analysis. Where relative amplification efficiencies were the same as match (+++++) was used; a twenty-fold reduction in relative efficiency is shown as +++; less than a 100-fold relative efficiency is shown as (+). For Huang *et al.,* 1992, the quantitative reduction in *Taq* polymerase extension of mismatches primer: template duplexes are represented by (+) when primer extension was reduced by less than10^{-6} fold; (++) represents a reduction of 10^{-6} fold; (+++) represents 10^{-3} to 10^{-5} fold; (++++) represents less than10^{-2} fold; and (+++++) represents the equivalent level of amplification as a match. Allele-specific PCR was used, followed by gel electrophoresis with subsequent ethidium bromide staining, followed by autoradiography for quantifying band intensities. For Bottema and Sommer, 1993, displayed data that exhibited reliable discrimination in all SNP combinations using allele-specific PCR, followed by electrophoresis and ethidium bromide staining. Significant optimisation was used, and the relative levels of amplification were not disclosed. For D'yachenko *et al.,* 1994, transversion mispairs exhibited lower amplification efficiency (+) than transitions (+++). Allele-specific PCR was used, followed by gel electrophoresis with subsequent ethidium bromide staining, followed by autoradiography for quantifying band intensities. For Day *et al.,* 1999, if product was detected by ethidium bromide staining after 10 cycles of allele-specific amplification, it is shown as (+++++); after 20 cycles (+++); and after 30 cycles (+). For Ayyadevara *et al.,* 2000, a complex representation of percentage amplification was given for two different template sequences. For clarity, these have been represented as discriminatory conditions (-) or not discriminatory (+++++). Allele-specific PCR was used, followed by gel electrophoresis with subsequent ethidium bromide staining, followed by autoradiography for quantifying band intensities.

Template base	*Primer base*	*Mismatch Type*	*Newton et al., 1989*	*Kwok et al., 1990*	*Huang et al., 1992*	*Bottema et al., 1993*	*D'yachenko et al., 1994*	*Day et al., 1999*	*Ayyadevara et al., 2000*
C	G	Match	+++++	+++++	+++++	+++++	+++++	+++++	+++++
	C	Transversion	NC	+	+	-	+	+++	-
	A	Transition	+++++	+++++	+++	-	+++	+++++	+++++
	T	Transversion	NC	+++++	++	-	+	+++	-
G	G	Transversion	NC	+++++	+	-	+	+++	-
	C	Match	+++++	+++++	+++++	+++++	+++++	+++++	+++++
	A	Transversion	NC	+	+	-	+	+	+++++
	T	Transition	+++++	+++++	+++	-	++++	+++	-
T	G	Transition	+++++	+++++	+++	-	+++	+++++	-
	C	Transversion	-	+++++	++++	-	+	+++	-
	A	Match	+++++	+++++	+++++	+++++	+++++	+++++	+++++
	T	Transversion	-	+++++	++	-	+	+++	-
A	G	Transversion	NC	+	+	-	+	+++	-
	C	Transition	+++++	+++++	+++	-	+++	+++	-
	A	Transversion	-	+++	++	-	+	+	+++++
	T	Match	+++++	+++++	+++++	+++++	+++++	+++++	+++++

Parameter		*Effect on reaction specificity in allele-specific PCR tests*
Core component concentrations		
Bottema and Sommer, 1993	Magnesium, dNTPs, primers, *Taq* DNA polymerase, template.	Lowering the concentration of all core components in the test can sometimes increase specificity. Too low and the yield will be adversely affected, too high and spurious amplification will occur. Excess magnesium will aid non-specific primer annealing, excess dNTPs will aid non-specific extension, and excess primers or enzyme will aid non-specific amplification.
Bottema and Sommer, 1993	Template	A starting template concentration that is too high can inhibit the PCR reaction. Too low, and primers will anneal to each other and form low molecular weight non-specific products by amplifying from themselves (primer dimers).
Cycling parameters		
Rychlik et al., 1990	Annealing temperature	If the annealing temperature is too low, non-specific DNA fragments are amplified since the primers will anneal in random areas of the template. Too high and the yield of the desired product and sometimes the purity is reduced due to poor annealing of primers. It is generally recommended to set the annealing temperature to 5°C above the T_m of the primers.
Sommer et al., 1992	Number of cycles	Decreasing the number of cycles may reduce the detection of any minor amplification of the mismatched allele. However, this does not usually affect the specificity.
Oligonucleotide design		
Rychlik, 1995; Wu et al., 1991	Length and stability	Long primers have greater stability and therefore mismatches at the 3' end are not destabilising. Shorter primers are recommended in AS-PCR. For optimal specificity, the 5' portion of the primer should be G/C rich and the 3' end should be A/T rich.
Sommer et al., 1992	Position of mismatch	Mismatches at the 3' end of the primer can disrupt extension since *Taq* initiates DNA synthesis from there. Additional mismatches can be incorporated at the 3' end, with most affect in the last three bases. Mismatches at the 5' end of a primer exert little/no effect on specificity.
Internal controls		
Bottema and Sommer, 1993	Spurious	A high magnesium concentration can be used to promote spurious fragments that can serve as internal controls. Here, this is not advised since this is likely to disrupt the specificity where it is truly required. Plus the control band is not defined and constant.
Newton et al., 1989	Additional primer set	Provides a constant and defined internal control fragment, and improves specificity by providing another substrate for *Taq* polymerase.
Instrumentation		
Wittwer et al., 1993	Transition times/ heating mechanism	Slow transition times often require longer times for PCR phase, increasing the opportunity for mismatch extension. Rapid cycling shortens the time for all phases of the PCR test, in particular, short extension times can prevent mismatch extension by decreasing opportunity for a mismatched primer annealed at the SNP site to be extended.

Table 3.3: Parameters for optimisation in AS-PCR tests. This list is not intended to be exhaustive.

PCR enhancer	*Effect on reaction specificity in allele-specific PCR tests*
Dimethyl sulfoxide (DMSO) *Shen et al., 1992*	Most commonly used PCR enhancer. Disrupts base pairing allowing complex secondary structure to be removed, which may normally prevent amplification. Good for GC rich templates. May affect the thermal activity if Taq polymerase and/or the extent to which strand separation of the amplified product is achieved at a particular temperature. Highly toxic substance.
Glycerol *Lu et al., 1993*	Reported to work in a similar way as DMSO by preventing the formation of secondary structure during amplification, and affecting the thermal profile of the enzyme. Glycerol more suitable for PCR enhancement that DMSO since it is not toxic.
Bovine Serum Albumin (BSA) *Kreader, 1996*	Thought to be useful in PCR tests where inhibitors may be present. Possibly protects *Taq* DNA polymerase from enzyme inhibitors.
N,N,N-trimethyglycine (Betaine) *Frackman et al., 2001*	Increase yield and specificity. Betaine is an isostabilising agent, equalising the contribution of GC- and AT- base pairing to the stability of the duplex. Therefore, useful in amplifying from templates which are GC-rich.
Tetramethylammonium chloride (TMAC) *Kovarova and Daraber, 2000*	Used to eliminate non-specific priming and improve the stringency of hybridisation reactions.
Additive effects *Frackman et al., 2001, Nagai et al., 1998, Waterfall and Cobb (unpublished data)*	PCR enhancers can often be used in combination and may exhibit additive effects. However, certain combinations and sub-optimal concentrations can completely inhibit the PCR. One such combination is glycerol and BSA.

Table 3.4: PCR enhancers used for improving the specificity and yield of PCR tests. This list is not intended to be exhaustive.

HotStart Method	*Basic Concept*
Manual *Li et al., 1990*	A reagent essential for amplification (i.e. Taq or $MgCl_2$) is withheld until the reaction temperature is sufficient to allow specific annealing under stringent conditions (>55°C). Can become tedious with large reaction sets since tubes need re-opening to add component. This also increases opportunity of contamination.
Antibody-based *Kellogg et al., 1994*	A neutralising monoclonal antibody to Taq DNA polymerase specifically blocks the enzyme activity until the reaction components are heated. This denatures the antibody such that it can no longer bind. Inhibitory properties become permanently inactivated upon heating.
Aptamer-based *Lin et al., 1997*	Aptamers are short single stranded DNA sequences that recognise *Taq* DNA polymerase with high affinity. They efficiently bind and inhibit the enzyme at lower temperatures (<40°C), but can be removed by heating, rendering the enzyme in its active form. The inhibitory effect is reversible.
Modified enzyme *Birch et al., 1996*	Uses a modified forms of Taq DNA polymerase e.g. AmpliTaq Gold™, which is inactive at low temperatures. Inhibitory properties become permanently inactivated upon heating and a drop in pH.
Magnesium precipitate *Barnes and Rowlyk, 2002*	Uses a novel buffer composition and reaction assembly protocol for PCR, including magnesium and phosphate combined at high concentrations in addition to standard buffer reagents. The magnesium containing precipitate is unavailable to DNA polymerase until thermal cycling occurs.
Double-stranded DNA fragment *Kainz et al., 2000*	Short double stranded DNA fragments are included in the reaction to inhibit Taq DNA polymerase. The inhibition is not sequence specific, but exclusively dependent upon the melting temperature of the fragments.

Table 3.5: Common methods of implementing Hot Start PCR. The list is not intended to be exhaustive.

These variable parameters should be optimised to accommodate the two non-variable parameters for each SNP assayed, i.e. surrounding DNA sequence and base composition of the SNP site. SNPs can fall into two subsets depending on their base composition, namely tranversions and transitions; tranversions represent the substitution of a purine (A or G) with a purine, or a pyrimidine (T or C) with a pyrimidine, whilst transitions represent substitutions of a purine with a pyrimidine or a pyrimidine with a purine (Brookes, 1999). A transition mispair maintains a chemical structure similar to the matched base pair since the same number of hydrogen bonds takes part in base pairing. In transversion mispairs however, the hydrogen bond complement is incomplete, creating a less favourable substrate for *Taq* DNA polymerase (Goodman *et al.,* 1993). It follows that a transversion SNP is generally more refractory to extension than a transition.

An obvious criterion that is often overlooked during optimisation of AS-PCR is the component interactions between core PCR reagents themselves (i.e. $MgCl_2$, dNTPs, *Taq* DNA polymerase, and primers) and their effects on mismatch extension. Literature suggests that a magnesium and oligonucleotide titration (in combination with additional mismatches in the allele-specific primer) is sufficient to obtain discriminatory conditions, yet Table 3.2 would suggest that this is not the case for all SNP combinations. Designing and testing primers of different lengths and sequence can involve large and expensive reaction sets, which can be avoided by considering the complex component interactions in a PCR test (Figure 3.2).

$MgCl_2$ and dNTPs have been shown to affect the efficiency of priming and extension by altering the kinetics of association and dissociation of primer-template duplexes at annealing and extension temperatures (Wolff *et al.,* 1993). These components also alter the efficiency with which *Taq* DNA polymerase recognises and extends such duplexes (Kwok *et al.,* 1990; Huang *et al.,* 1992). Of particular relevance to AS-PCR methods, the concentration of $MgCl_2$ and dNTPs required depends largely on the target and primer sequences. The presence of excess magnesium may promote non-specific amplification products and insufficient concentrations may reduce product yield (Rychlik, 1990).

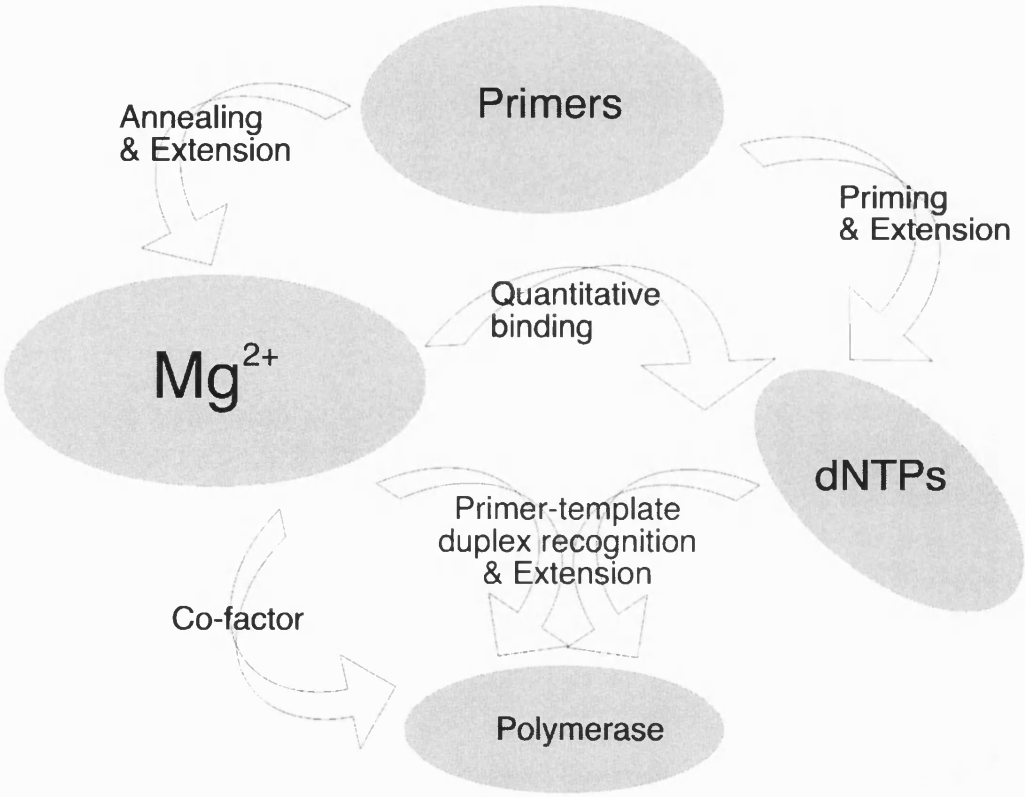


Figure 3.2: Diagrammatic representation of key reaction component interactions in a PCR test.

Also, dNTPs quantitatively bind Mg^{2+}, so that any modification of dNTP concentration will require a compensatory adjustment of $MgCl_2$ (Innis, 1990). *Taq* polymerase requires free Mg^{2+} as a cofactor so any excess of dNTPS can have a detrimental effect on product yield by chelating cofactor ions (Linz *et al.,* 1990).

This chapter describes the use of a multifactorial optimisation strategy to define conditions allowing SNP discrimination regardless of base composition. The method is based on a process widely used for development trials during industrial process design where the aim is to generate enough information to establish optimal conditions for a particular process using the minimal number of experiments possible (Taguchi, 1986). In 1994, modified Taguchi methods were applied to optimising standard PCR, defining component concentrations that generate maximal yield and minimal spurious amplification. For standard PCR, 'control factors' are selected as the components most likely to exert greatest effect on the outcome of the test. Taguchi then uses a number of progressive trials to predict a combination of component concentrations that will lead to optimal performance (Cobb and Clarkson, 1994). If these results are satisfactory then further experiments are unnecessary. Further improvement may be achieved with subsequent trials using a narrower range of concentrations for each component.

Here, the Taguchi format is used to establish discriminatory conditions for SNP genotyping in a small number of reaction sets. The Taguchi method also facilitates an upstream regression analysis used here to study the effects of individual PCR components on amplification.

3.1 Results

3.1.1 Conventional allele-specific PCR tests

A schematic of conventional allele-specific PCR is shown in Figure 3.1. Primers WT_CP_374 and ASA_1X (where X is G, C, A or T) were designed to direct the amplification of a 374 bp fragment from a β-actin gene segment inserted into pGEM-T Easy Vector-based plasmid DNA constructs (Section 2.1). Nine PCR reactions were performed for each optimisation set according to the general protocol described in Section 2.0.1, except the four components most likely to affect the outcome of the PCR (primers, *Taq* DNA polymerase, $MgCl_2$ and dNTPs) were aliquoted according to an orthogonal array (Table 3.6; Cobb and Clarkson, 1994), at three different concentrations (Table 3.7). Internal control primers ICF and ICR (0.04 μM each) were added to amplify a 223 bp control fragment. Sixteen sets were performed in total, including the four matched PCR tests, and twelve possible mismatched tests.

PCR products were analysed by agarose gel (1.5%) electrophoresis and maximal optical density (OD_{max}) readings were obtained (Section 2.4.1). Gel images and corresponding OD_{max} data are presented in Figure 3.3 and Table 3.8, respectively. Amplicon yields varied within sets according to component concentrations, and between sets according to base complementation. Yields were greater in perfectly matched than mismatched tests for all templates. The control fragment was present in all PCR tests.

A qualitative analysis of the amplicon yields allowed the selection of reaction conditions that amplified product in the presence of a matched nucleotide, yet under identical conditions, remained refractory in the presence of a mismatch. Conditions were elucidated for all SNP combinations and used in subsequent genotyping tests to confirm the optimal nature. All twelve polymorphisms were genotyped in triplicate (Figure 3.4). For each SNP combination, the conditions enabled amplification of a matched template, with little or no amplification in the presence of a mismatch.

Reaction	*[1]*	*[2]*	*[3]*	*[4]*
1	A	A	A	A
2	A	B	B	B
3	A	C	C	C
4	B	A	B	C
5	B	B	C	A
6	B	C	A	B
7	C	A	C	B
8	C	B	A	C
9	C	C	B	A

Table 3.6: Orthogonal array for testing four variables each at three levels. Variables [1], [2], [3] and [4] are primer set, *Taq* DNA polymerase, $MgCl_2$ and dNTPs, respectively.

Level	*A*	*B*	*C*
BFR/ASA1X (μM)	0.2	0.4	0.6
Taq (U)	0.5	1.25	2.5
$MgCl_2$ (mM)	0.5	1	1.5
dNTPs (μM)	80	120	200

Table 3.7: Concentration levels for components used in orthogonal array (Table 3.6).

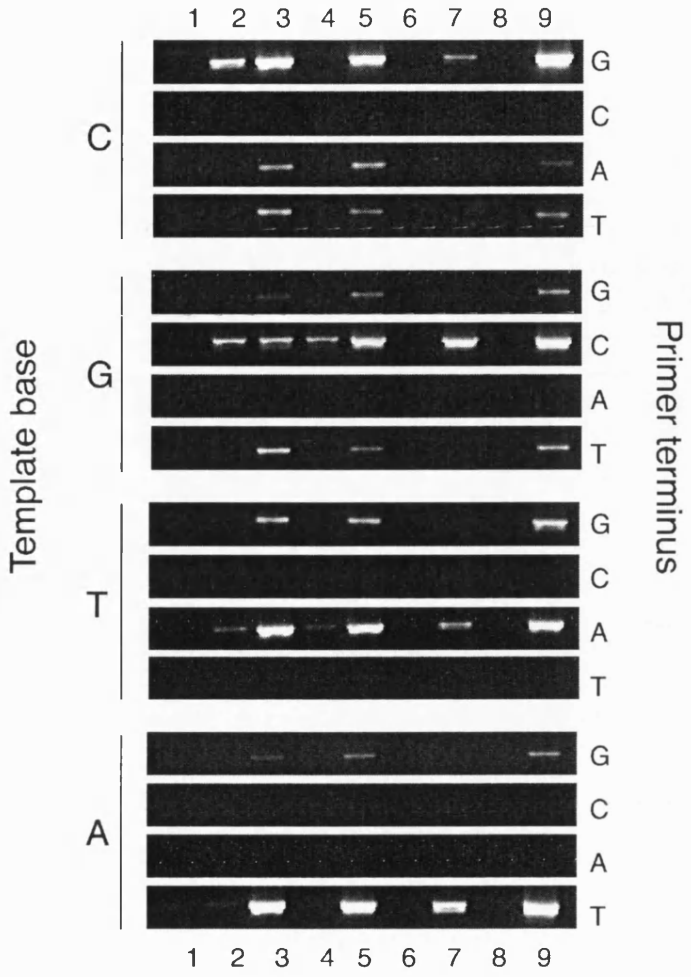


Figure 3.3: Typical gel images from optimisation tests covering all possible matched and mismatched base complementations. In each case, the DNA band is 374bp. No spurious amplification was observed across the reaction set. Corresponding OD_{max} data is displayed in Table 3.8.

Tenplate: primer	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
C:G	**-**	**107.1**	**120.3**	**1**	**110.6**	**-**	**52.3**	**-**	**119.3**
C:C	-	-	-	-	-	-	-	-	-
C:A	-	-	58.3	-	72.5	-	-	-	23.5
C:T	-	-	79.1	-	39.3	-	-	-	47.5
G:G	-	-	-	-	39.8	-	-	-	68.12
G:C	**-**	**100.0**	**99.4**	**71.6**	**109.0**	**-**	**110.8**	**-**	**111.5**
G:A	-	-	-	-	-	-	-	-	-
G:T	-	-	95.5	-	42.0	-	-	-	77.3
T:G	-	-	81.1	-	85.0	-	-	-	101.5
T:C	-	-	-	-	-	-	-	-	-
T:A	**-**	**34.2**	**113.6**	**23.3**	**109.7**	**-**	**58.4**	**-**	**110.2**
T:T	-	-	-	-	-	-	-	-	-
A:G	-	-	25.9	-	42.0	-	-	-	53.7
A:C	-	-	-	-	-	-	-	-	-
A:A	-	-	-	-	-	-	-	-	-
A:T	**-**	**-**	**131.4**	**-**	**127.6**	**-**	**118.3**	**-**	**129.3**

Table 3.8: Product yields from allele-specific PCR tests covering all base complementations; represented as maximal optical density (OD_{max}) values. Numbers in **bold** type represent matched amplifications. A dash (-) represents no amplification. Corresponding gel image data is displayed in Figure 3.3.

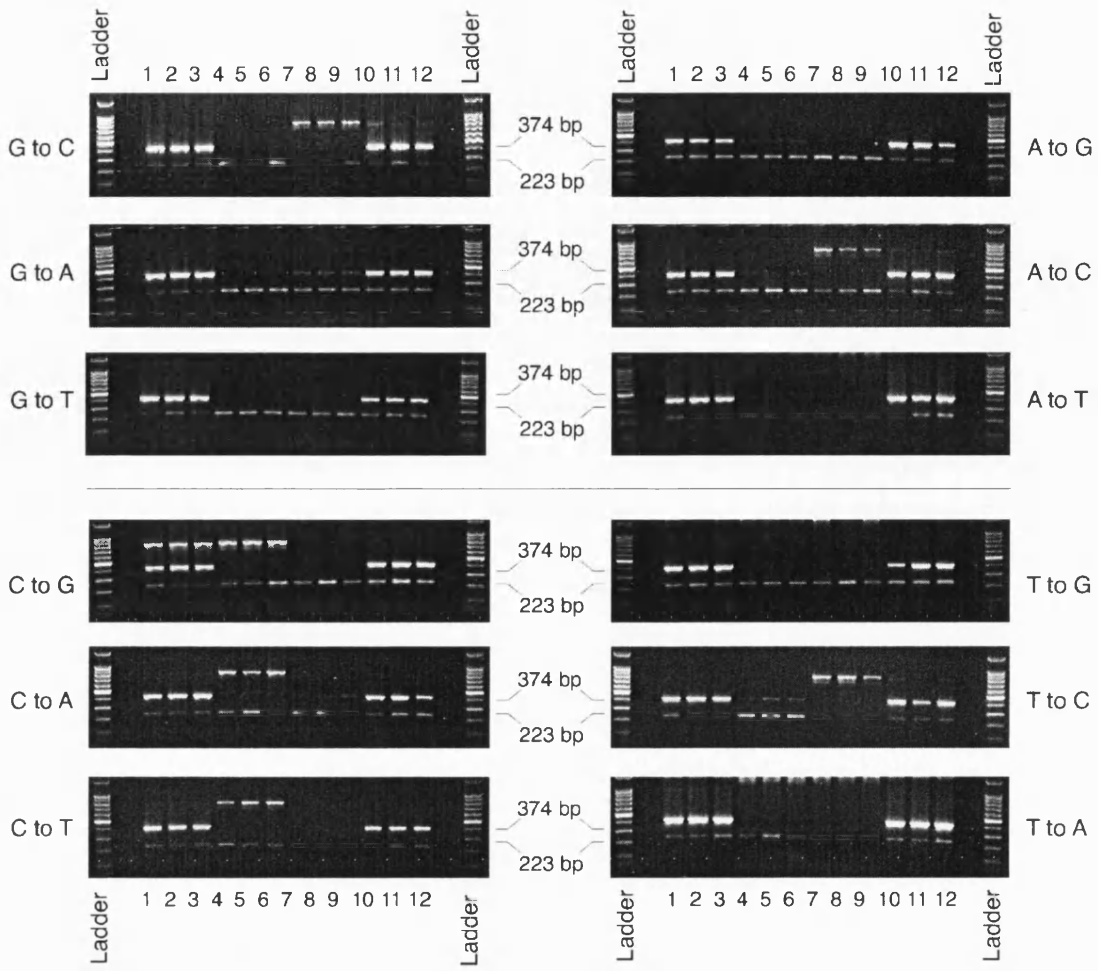


Figure 3.4: Typical gel images displaying triplicated reactions performed using the discriminatory conditions defined by the optimisation strategy. The target region is 374 bp in each case. Wells 1-3 contain wild type template with wild type primer amplifications; wells 4-6 represent wild type template with mutant primer; wells 7-9 represent mutant template with wild type primer; and wells 10-12 represent mutant template with mutant primer. The internal control fragment (223 bp) was present across the reaction set. Conditions defined by Reaction 7 were used to detect SNPs G→C, G→A, G→T, C→A, C→T, A→G, A→C and T→C; Reaction 2 was used to detect C→G; Reaction 5 was used for A→T and T→G; and Reaction 9 was used for detecting a T→A polymorphism.

In certain cases, artefactual amplification was observed (experiments detecting SNPs G→C, A→C, C→G, C→A, T→C, and C→T). The size of this artefactual fragment is consistently ~1000 bp, which would be the approximate size of the fragment, formed by primers ICF and BFR (Appendix I). Alternatively, since a common factor in the experiments where this artefact occurs is the detection of base 'C', it is possible that the allele-specific primer has complementarities with another site on the plasmid sequence. This did not interfere with overall specificity.

3.1.2 Single tube genotyping of sickle cell anaemia using bi-directional PCR

A schematic of bi-directional AS-PCR is shown in Figure 3.5. Two allele-specific primers were designed with complementarities to normal β-globin (SC_WT_AS) or sickle cell (SC_MUT_AS) genes. These were arranged in a bi-directional orientation such that both primers terminated on the SNP site. Primer SC_WT_AS has a single 'A' nucleotide at the 3' end, matching the 'T' nucleotide on the non-coding strand of the β-globin gene. Primer SC_MUT_AS has a single 'A' nucleotide at the 3' end matching the 'T' base on the coding strand of the sickle cell gene. The wild type outer primer (WT_OP_517) was positioned 517 bp downstream on the opposite strand of SC_WT_AS. The mutant outer primer (MUT_OP_267) was placed 267 bp upstream of SC_MUT_AS. Internal control amplification was not required since any successful PCR will generate an allele-specific amplicon.

For optimisation, four critical components ([1] WT primer set, [2] MUT primer set, [3] $MgCl_2$ and [4] dNTPs) were selected and aliquoted according to Table 3.3 with final concentrations according to Table 3.9. Two identical optimisation sets were run containing ~10 ng of [A] wild type or [B] mutant control DNA (Figure 3.6/Table 3.7). Other reaction components included Mg-free PCR buffer, 0.025 $U.\mu l^{-1}$ *Taq* DNA polymerase with HS TaQUANT-OFF™ (Q-Biogene) and PCR-grade water to a total volume of 25μl. Cycling conditions and product analysis were identical to those described in Section 3.1.1. Product yield varied considerably across the reaction sets according to varying component concentrations, and between reaction sets and according to template zygosity. The wild type 517 bp amplicon only was amplified in set [A], and the mutant 267 bp fragment only was amplified in set [B].

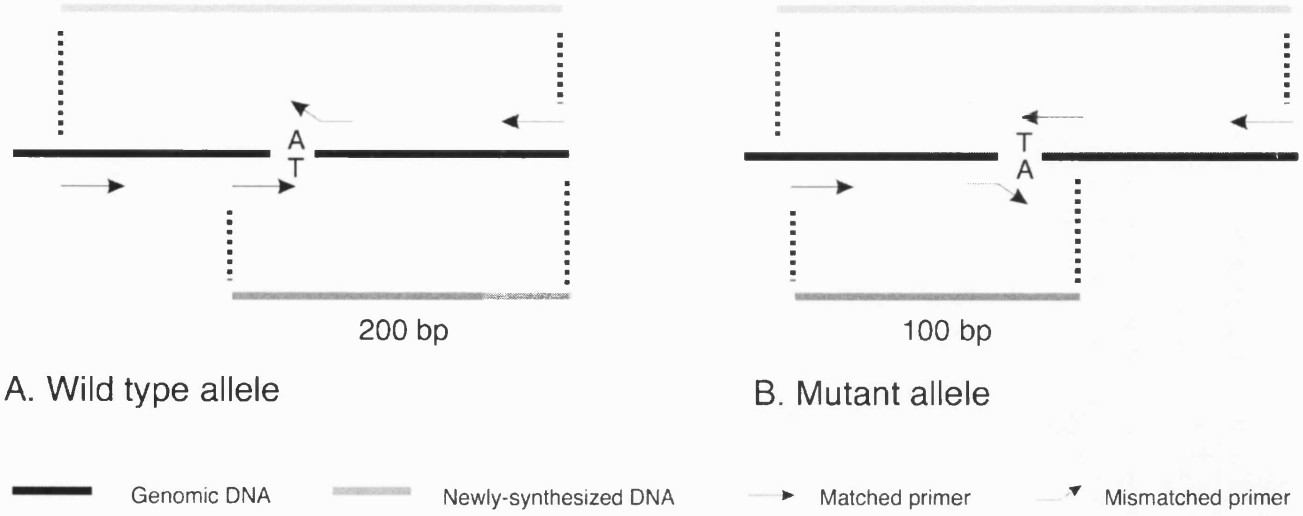


Figure 3.5: Schematic representation of bi-directional allele-specific PCR. Three possible fragments can be formed depending upon substrate DNA zygosity, as well as enzyme system and cycling parameters. The 200 bp amplicon defines the allele-specific product formed when wild type DNA is present, and only wild type primers are extended. The 100 bp amplicon defines the allele-specific product formed when mutant DNA is present and only the mutant primers are extended. A heterozygote is identified by the amplification of both PCR products. The target region (300 bp) amplified by non-allele specific outer primers is shown in pale grey. The enzyme system selected was inefficient at amplifying longer sequences and in the presence of competition from a smaller fragment insufficient target region was amplified for visualisation on a stained gel.

Level	*A*	*B*	*C*
SC-WT-AS/ WT-OP517 (μM)	0.2	0.4	0.8
SC-MUT-AS/ MUT-OP267 (μM)	0.2	0.4	0.8
$MgCl_2$ (mM)	0.5	1.0	2.0
dNTPs (μM)	80	120	200

Table 3.9: Concentration levels for components used in orthogonal array for optimisation of bi-directional allele-specific PCR

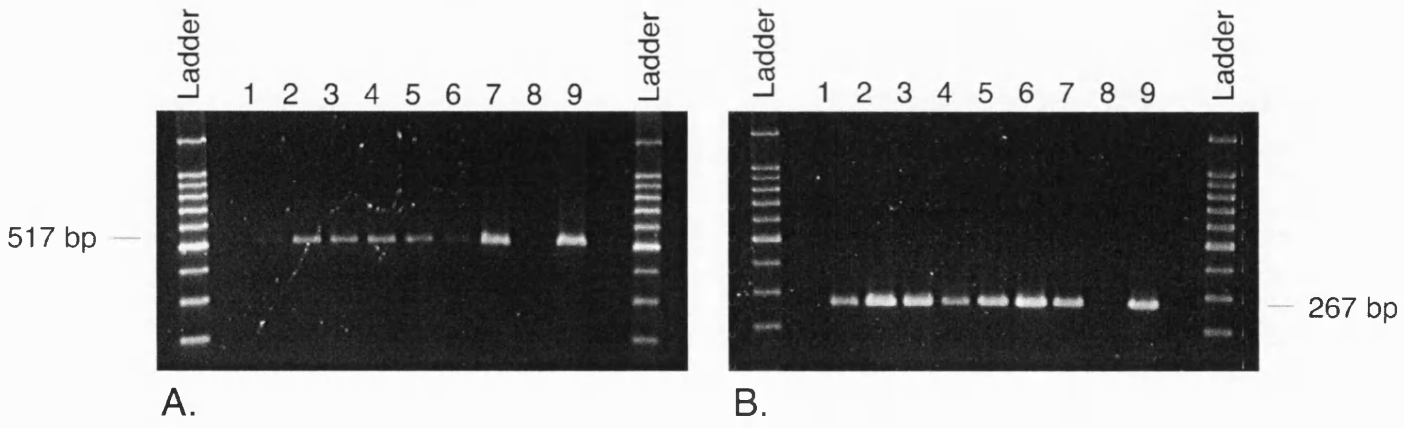


Figure 3.6: Typical gel image data of PCR optimisations. Outer lanes contain 100 bp ladders. Reactions 1 to 9 according to the orthogonal array (Table 3.6) are labelled accordingly. Set [A] represents reactions in which wild type template DNA was the substrate. Set [B] represents reactions in which mutant (HbS) template DNA was included. The 517 bp and 267 bp allele-specific fragments were generated in sets [A] and [B], respectively. Product yield varied across each reaction set. Table 3.10 represents with OD_{max} data of the amplicon bands, used as a representation of product yield/reaction score.

Rxn	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
[A] OD_{max}	5.13	90.83	80.30	63.40	63.70	-	132.27	-	160.67
[B] OD_{max}	113.6	172.53	169.20	108.73	161.13	183.67	144.80	-	165.93

Table 3.8: Product yield data (OD_{max}) from optimisation sets [A] and [B] in Figure 3.6. Samples giving no product are designated by (-).

Diagnostic conditions were selected by *qualitative* analysis of PCR yields from optimisation sets [A] and Set [B]. Optimal conditions for zygosity determination are those that amplify each allele-specific amplicon at similar efficiencies, measured by the intensity of specific bands on the gel. In this study, the conditions and interactions of Reaction 9 (Tables 3.6 and 3.9) satisfied these criteria, producing 517 bp and 267 bp bands with OD_{max} values of 160.67 units and 165.93 units respectively. A small trial was performed using these conditions in which 14 human genomic DNA samples of unknown zygosity were used as template using considered optimal discriminatory conditions (Figure 3.7). The initial DNA concentration of each sample ranged from 5 to 25 $ng.\mu l^{-1}$. Reactions generating the 517 bp fragment only identified a normal genotype. The 267 bp fragment was amplified confirming a homozygous mutation and sickle cell carriers (HbA/S) were identified by the amplification of both amplicons. Genotypes were 100% concordant with those generated by the source laboratory.

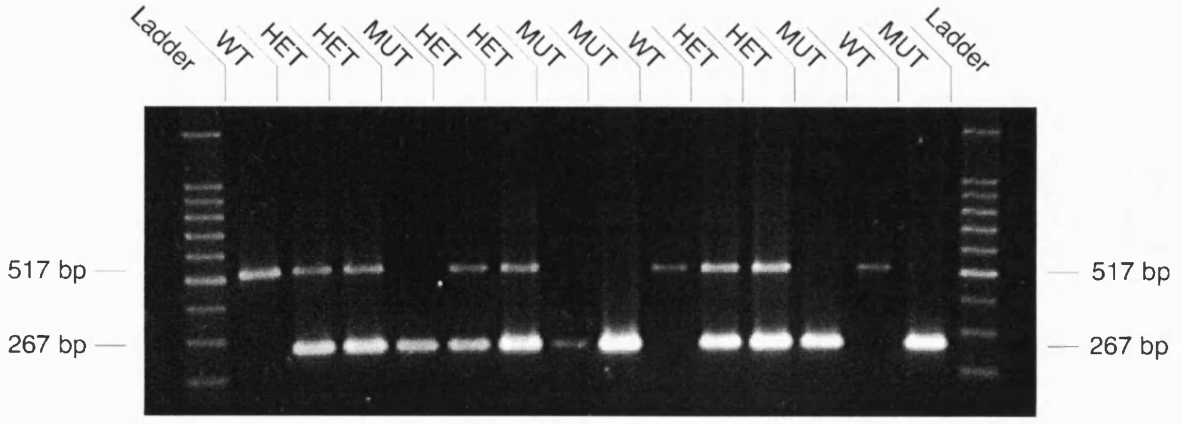


Figure 3.7: Gel image data from a small-blinded trial. Fourteen human DNA samples of varying zygosity were selected and amplified under conditions specified by Reaction 9 (Table 3.6/Table 3.9). Diagnosed template zygosity is shown on the gel image. Three normal samples were identified by the amplification of the 517 bp products only. Six heterozygous samples were identified by the amplification of both 267 bp and 517 bp fragments. For heterozygotes, the 267 bp band was slightly more intense than the 517 bp band, likely to be due to staining disparity between optimisation sets [A] and [B]. Five homozygous sickle cell samples were identified by the amplification of the 267 bp amplicon only.

3.1.3 *Reaction components affecting product yield (SN_L analysis)*

Amplicon yields from optimisation sets A and B were "scored" by obtaining OD_{max} data from gel images (Table 3.10). Those giving non-allele specific fragments or no amplification at all were given a score of 1. The 'yield scores' were used in a simple downstream mathematical analysis to determine which components exerted maximal effect on product yield at a particular concentration. Generally, the theoretical target of the PCR test in the presence of a matched template is to increase the product yield so that it is as large as possible. Under these circumstances, the Taguchi method recommends the quadratic loss function,

$$SN_L = -10\log_{10}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right) \qquad (3.1)$$

assuming that variable, *y*, denotes the product yield measured for a particular combination of factor settings in a given experiment (total of *n* repeated measurements per experiment). The signal-to-noise ratios (SN_L) for each parameter at each concentration level tested (Table 3.11) were plotted versus concentration for each component, and refined by polynomial regression to obtain curves whose maximum represent the reaction optima (Figure 3.8; Cobb and Clarkson, 1994).

Variable	*Level- Gel A*			*Level- Gel B*		
	A	**B**	**C**	**A**	**B**	**C**
WT primers	18.94	4.77	4.77	43.12	42.93	4.77
MUT primers	18.95	4.77	4.77	41.01	4.77	44.73
$MgCl_2$	1.68	39.94	38.12	4.77	42.88	43.94
dNTPs	18.94	4.77	4.77	42.94	44.32	4.77

Table 3.9: SN_L values calculated (using equation 3.1) using product yield data from optimisation sets A and B (Table 3.10) as reaction 'scores'.

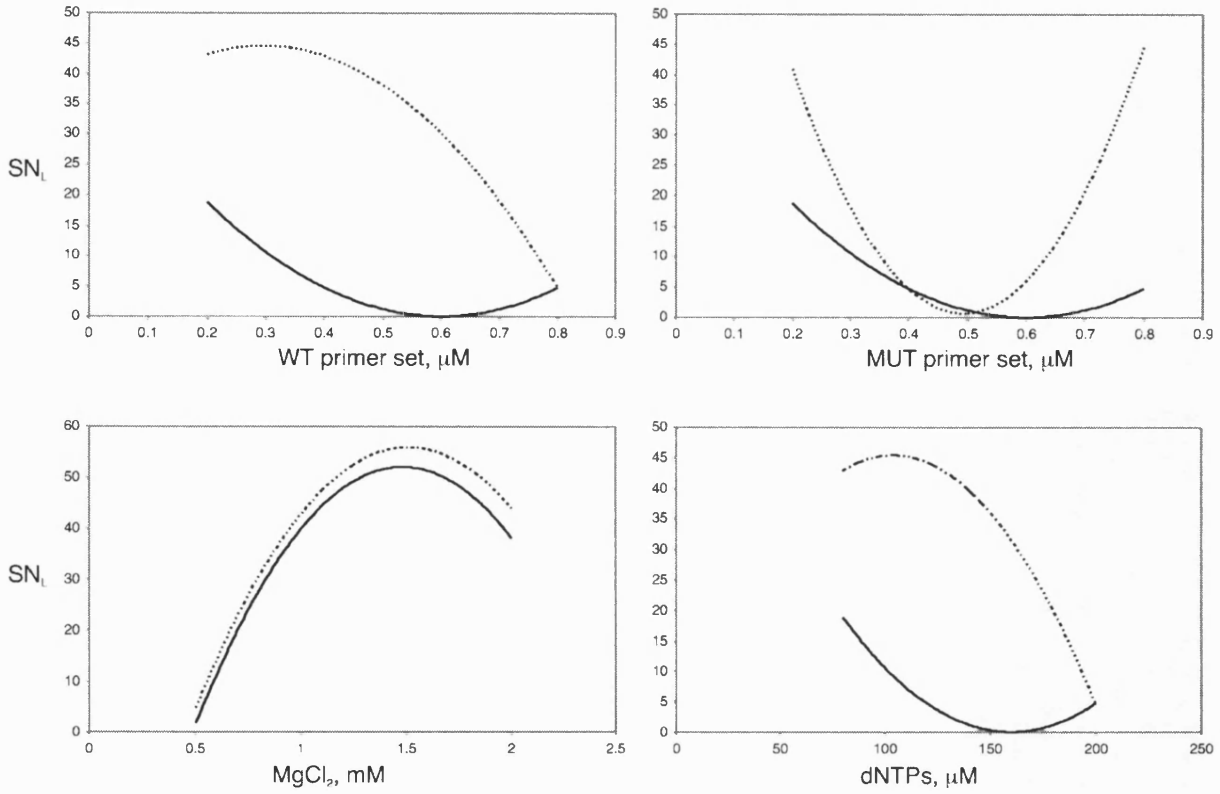


Figure 3.8: Effects of reaction components on the SN_L ratios for the amplification of a wild type 517 bp amplicon from wild type DNA (solid line); and a mutant 267 bp amplicon from HbS DNA (broken line). Increased SN_L value represents a larger effect on product yield than a low SN_L value.

3.2 Discussion

In AS-PCR methods, reaction conditions are identified that give a good yield for matched genotypes, yet under identical conditions give no amplification in the presence of a mismatch. Importantly, both products should be amplified in the presence of heterozygous DNA. The traditional methods of achieving high discrimination were discussed previously, the most common of which is to introduce an additional mismatch to the 3' end of the allele-specific primer. Using this approach, multiple primers may require design and testing, because the nature of the base composition of the additional mismatch and surrounding sequence can sometimes render the primer refractory to extension, even in the presence of a matched template. This can prove costly and lead to large reaction sets, especially if combined with a sequential magnesium and oligonucleotide titration, and annealing temperature optimisation.

Here, it is shown that an optimisation based on the core PCR component interactions can quickly establish discriminatory conditions in a small trial, without the need for additional mismatches and design considerations. Conventionally, a sequential investigation of the effects and interactions of four reaction components each at three concentration levels would require an experiment with 81 (i.e. 3^4) separate reactions. Using the Taguchi method the same trial was carried out in just nine reactions with wild type and mutant control templates (Cobb and Clarkson, 1994).

The core components considered most likely to affect the outcome of the PCR were selected and varied according to an orthogonal array. Where 3 concentration levels are tested for each reaction component, the number of experiments (E) is calculated from the equation $E = 2k + 1$, where k is the number of factors to be tested (Cobb and Clarkson, 1994). Therefore, just nine reactions per optimisation set were required to test four components. For conventional allele-specific PCR, these were primers, magnesium chloride, dNTPs, and *Taq* DNA polymerase. For bi-directional PCR, these were magnesium chloride, dNTPs, wild type and mutant

primer sets. The primer concentrations within each set were equimolar since there is no evidence to suggest that using a different ratio of forward and reverse primers has a beneficial effect in PCR (Henegariu *et al.,* 1997).

The optimisation strategy was applied to conventional AS-PCR where discriminatory conditions were established for the detection of all SNP combinations using a pGEM-T Easy Vector plasmid template with a single variant nucleotide (Figure 3.3), and verified by running triplicated tests (Figure 3.4). This shows a significant enhancement to AS-PCR methods when compared to the data in Table 3.2.

In addition, a gel-based single tube adaptation of allele-specific PCR was developed (Figure 3.5), where allele-specific amplicons were identified by size discrimination. The test was combined with the multifactorial optimisation strategy, taking into account the core component interactions in the PCR, and additional mismatches or PCR enhancers were not required to achieve discrimination (Figure 3.6). Conditions were quickly established for the detection of an A→T polymorphism causing sickle cell anaemia by taking into account the fact that at equimolar concentrations, amplification by the two primer sets might be weak or strong at a given locus. This event is dependent upon primer stabilities, binding efficiencies, priming competition, product size and concentration of other reaction components. Unique to bi-directional PCR, each primer is complementary to a different part of the DNA sequence, introducing additional variability to the reaction.

The optimal nature of the conditions was subsequently verified in a small-blinded trial where results were 100% concordant with RFLP data from the source laboratory (Figure 3.7). Importantly, these conditions discriminated the transversion A: A and T: T mismatches designated by the sickle cell polymorphism which was not the case in earlier reports (Kwok *et al.,* 1990; Huang *et al.,* 1992). However, conflicting studies were performed using different template sequences and reaction conditions, introducing variability.

An informative downstream application of Taguchi optimisation is the ability to quantitatively study the biochemical basis of component interactions and priming competition using regression analysis. Product yield data from optimisation sets was used to satisfy a quadratic equation where the theoretical target is to generate the highest amplicon yield possible. The specific SN_L ratios mathematically penalise small deviations from the theoretical target, generating a value that is the highest possible average with the lowest standard deviation (Table 3.9). Optimal conditions are those giving the largest SN_L value (Cobb and Clarkson, 1994).

The SN_L analysis using equation 3.1 was applied to bi-directional PCR, revealing that the optimal magnesium requirement for the generation of maximal product yield and specificity was 1.5 mM irrespective of template zygosity. Likewise, the optimal concentration of dNTPs in sets A and B were similar, with lower concentrations being favoured. Excess dNTPs may chelate the magnesium co-factor ions of *Taq* decreasing processivity, or reduce yield and specificity.

The regression profiles of primer sets generated by wild type and mutant DNA samples demonstrated variation in primer stabilities, efficiencies and competition (Figure 3.8). For wild type DNA the aim is to direct amplification of the 517 bp fragments only. The lower concentration of the wild type primer set tested gave the largest SN_L value relating to product yield. Intermediate concentrations decreased the SN_L, and the upper concentration showed an incremental increase in SN_L. An identical profile was observed with the mutant primer set, optimally redundant in the presence of wild type DNA. It is possible in the presence of low primer set concentrations, i.e. conditions of minimal competition, wild type primers can direct amplification of the 517 bp fragment. When the mutant allele is present, the aim is to direct amplification of the 267 bp amplicon only. The optimal concentration of the mutant primer set was predictably at the upper level. The inclusion of lower concentrations of wild type primers allowed maximal amplification of the mutant fragment, wherein the wild type set was exhibiting least competition, demonstrating the significant role of primer interactions in bi-directional amplifications.

In accordance with previous experiments (data not shown), the conditions selected for the blinded trial (0.8 μM each primer, 1 mM $MgCl_2$ and 80 μM dNTPs) were not in accordance with optimal component levels calculated by SN_L values, due to the different objectives of each enquiry. The theoretical target is to define conditions amplifying a maximum yield. This differs from the practical target of specifying conditions where each allele is specifically amplified with equal efficiency. In order to achieve this, sub-optimal conditions for the amplification of one allele may be combined with optimal conditions for amplification directed by a less efficient primer set.

Bi-directional amplifications were performed using an antibody-based Hot-Start *Taq* DNA polymerase. The enzyme demonstrated high fidelity, only amplifying allele-specific products with minimal spurious artefacts. Additionally, the enzyme was inefficient at amplifying larger fragments (>600 bp) when in competition with a smaller fragment, thus the non-allele specific fragment spanning the whole target region was not amplified. In similar studies, this larger fragment has been used as an internal control to avoid the incorporation of a distal primer pair conventionally required for diagnostic tests (Liu *et al.,* 1997). However, bi-directional PCR reaction will always yield a fragment, serving as an internal positive control for the PCR. Here, the processivity of *Taq* directed toward generating allele-specific products only, generating high amplicon yields without spurious amplification.

Testing the effects of four reaction variables simultaneously reduced the time taken to optimise the test, lowering the associated reagent costs. When combined with the single-tube bi-directional assay, twice as many samples can be genotyped in a single PCR run compared to other gel-bases analyses, addressing throughput issues. In both cases, additional mismatches in allele-specific primers were not required and a standard annealing temperature was used in all tests. This suggests that the need to examine the more costly variables, such as primers of different length and sequence, can be overridden by considering the classical interactions of core PCR components. If conditions are not established, alternative methods to improve discrimination can then be used. The optimisation strategy may also be applied to other molecular biology assays to establish optimal conditions according

to a theoretical target, including the effects of different additives on a PCR test, and combinations of additives (data not shown).

The utility of gel-based single tube adaptation of AS-PCR in clinical testing is recognised by very recent publications describing similar methods (Brightwell *et al.*, 2002). However, the rate-limiting step of these assays is the necessity for size discrimination of the resulting allele-specific PCR products using agarose gel electrophoresis. Chapter four describes the application of melt curve analysis and real time detection to bi-directional PCR, to provide rapid SNP genotyping.

This work resulted in a publication in Nucleic Acids Research, 2001 (Appendix III).

Chapter Four

Rapid SNP Genotyping using Fluorescent Bi-directional PCR and Melt Curve Analysis

4.0 Introduction

In 1992, Higuchi *et al.* published pioneering research in which the advent of 'real-time' PCR was presented. A dsDNA specific binding dye, ethidium bromide, was incorporated directly into a PCR reaction mixture to visualise DNA amplification in real time by directing excitation illumination through walls of the amplification vessel before or after, or even continuously, during thermocycling (Higuchi *et al.,* 1992). This work was pivotal in revealing the shape and characteristics of a PCR reaction curve (Figure 1.9), and increased the utility of PCR methods for clinical application, i.e. quantitative PCR and mutation detection.

This prompted the development of PCR instruments with integrated optics systems allowing continuous monitoring of PCR amplification in a closed reaction vessel (Table 4.1). Concurrently, a new class of dyes (unsymmetrical cyanine dyes) was synthesised including SYBR® Green I, PicoGreen, SYBR® Green II, YOYO®, YO-PRO®, and SYBR® Gold (Tuma *et al.,* 1999), which exhibit more favourable properties for DNA analysis than ethidium bromide. As a group, the unsymmetrical cyanine dyes are characterised as having low intrinsic fluorescence, large fluorescent enhancements upon binding to nucleic acids, and high fluorescence quantum yields when complexed with nucleic acids, producing large signal-to-background ratios. In combination with their apparently high binding affinities for nucleic acids, and greater sensitivity than ethidium bromide, these dyes are commonly included in PCR reaction mixes for real-time monitoring (Lay and Wittwer, 1997).

Double-strand specific DNA binding dyes also allow simultaneous product identification in the same closed reaction vessel by DNA melting curve analysis. Introduced in Chapter One, 'DNA melting' describes the denaturation of DNA from a double-stranded to single-stranded state (ssDNA). The specific temperature at which the DNA is 50% single stranded is known as the "DNA melting temperature" (T_m). The transition from dsDNA to ssDNA can be observed in real time by including a dsDNA specific reporter dye (e.g. SYBR Green I) into the PCR mixture, and continuously monitoring fluorescence as the temperature is slowly elevated (c.f. $0.1°C.s^{-1}$) from low (~60°C) to high temperatures (~95°C). A rapid loss of fluorescence is observed at the specific melting temperature of the DNA, characterised by the inability of the dsDNA specific dye to remain bound. The read-out from such monitoring is called a "melt curve" (Figure 1.10A, pg. 28), which is converted to a "melt peak" (or series of melt peaks) by plotting the first negative derivative of fluorescence loss, F, with respect to temperature, T ($-dF/dT$ vs. T; Figure 1.10B). Since the shape and position of a melt peak depends markedly on the GC/AT ratio, length and sequence of the amplicon, products of a similar size can be differentiated in a reaction mixture (Ririe *et al.,* 1997).

The LightCycler™ PCR instrument, used in these experiments, delivers an additional advantage of speed. Previously, the bi-directional PCR test was based on slow Peltier-effect heat block thermal cycling, the platform upon which most thermal cyclers are based (Table 4.1). Peltier elements operate essentially by applying a current through two conductors with dissimilar electron densities causing heat to be absorbed on one side and released on the other- by controlling the direction of electron flow, such a device can act as both a heating and cooling mechanism (Pierce, 2002). Using this mechanism, a typical DNA amplification may require in excess of 3 hours, although 2-hour amplifications are possible if adjustments to minimise sample denaturation and annealing times are made (Wittwer *et al.,* 1993). The LightCycler™ displays improved temperature control by alternating heated and ambient air as the medium for heat transfer. The low mass of air means that very rapid temperature exchange rates can be achieved within the thermal chamber ($20.0\ °C.s^{-1}$; Wittwer *et al.,* 1997). Also, glass-capillary tubes are used in place of plastic tubes with a large surface-to-volume ratio, further improving the thermal

properties for heat exchange. Consequently, 30 PCR cycles can be completed in less than 30 minutes using the LightCycler™ platform, delivering higher throughput.

The assay described in this chapter describes rapid cycle bi-directional PCR combined with melt curve analysis using SYBR Green I, to eliminate a major bottleneck in this type of diagnostic test, predominantly due to the relatively limited throughput of gel electrophoresis systems. The integrated system introduces homogeneity to a previously heterogeneous system, and a higher throughput suitable for a clinical environment. Detecting the three clinically relevant SNPs described in Section 1.1 validated the test.

Company	*Product*	*Excitation source*	*Excitation range, nm*	*Emission Range, nm*	*Detector*	*Changeable filters*	*Multiplex capabilities*	*Max. Sample capabilities*	*Heating system*
Applied Biosystems	7700 Sequence Detection System	Argon lamp	488	500-650	CCD	No	>4	96	Vapour and compression
Cepheid	SmartCycler	4 LEDs	450-650	510-750	Photodiodes	Yes	4	16	Ceramin heaters
BioRad	i-cycler	Tungsten-halogen lamp	400-700	400-700	10-bit CCD	Yes	4	96	Peltier and Joule
Corbett	Rotor-Gene 3000	4 LEDs (blue, green, orange, red)	470-625	510-660	PMT	Yes	4	36 or 72	Air-exchange
Roche	LightCycler	Blue LED	470	530, 640, 710	Photodiodes	No	3	32	Air-exchange
MJ	Opticon	LED	488	530	PMT	No	1	96	Peltier
MJ	Opticon II	LED	TBA	TBA	PMT	No	2	96	Peltier
Stratagene	Mx4000	Tungsten-halogen lamp	350-750	350-830	PMT	Yes	4	96	Peltier with connective and resistive technology

Table 4.1: Specifications of selected commercially available real-time thermocyclers with integrated optics systems. Data accumulated from Manufacturer's web sites.

4.1 Results

4.1.1 Assay design

Figure 4.1 shows a schematic representation of the fluorescent bi-directional PCR test. Primer sets were designed to detect three clinically relevant SNPs. The A→T mutation in codon 6 of the β-globin gene related to sickle cell disease was chosen as a SNP target. Primers SC_WT_AS and SC_WT_OP were designed to direct the amplification of an 86 bp product from the normal β-globin gene, with a calculated T_m of 87.0°C (wild type primer set). SC_WT_AS has an 'A' nucleotide at the 3' end matching the 'T' base on the coding strand of the normal β-globin gene. Primers SC_MUT_AS and SC_MUT_OP were designed to amplify a 112 bp fragment from the mutant DNA, with a calculated T_m of 85.0°C (mutant primer set). SC_MUT_AS has an 'A' nucleotide at the 3' end, matching the 'T' base on the coding strand of the sickle cell mutation. The sequence of interest and primer alignments is shown in Figure 4.2.

Secondly, primers were designed for the detection of the most common polymorphism in the *HFE* gene related to hereditary haemochromatosis (HH), namely the G→A polymorphism at nucleotide position 845 of the open reading frame (Ugozzoli *et al.*, 2002). The resulting HFE protein exhibits a change from cysteine to tyrosine at amino acid position 282, referred to as 'C282Y'. Primers C282Y_WT_AS and C282Y_WT_OP were designed to direct the amplification of a 76 bp product from the normal *HFE* gene, with a calculated T_m of 87.0°C (wild type primer set). C282Y_WT_AS has a 'G' nucleotide at the 3' end matching the 'C' base on the non-coding strand of the *HFE* gene. C282Y_MUT_AS and C282Y_MUT_OP amplify a 72 bp fragment from homozygous mutant DNA, with a calculated T_m of 84.0°C (mutant primer set). C282Y_MUT_AS has a 'T' nucleotide at the 3' end, matching the 'A' polymorphic base on the coding strand. The sequence of interest and primer alignments is shown in Figure 4.3.

Primers were also designed for detection of the less common missense C→G transversion in the *HFE* gene at nucleotide position 187 of the open reading frame, also linked to increased susceptibility to HH when inherited with the C282Y mutation (Ugozzoli *et al.,* 2002). This causes an amino acid change in the HFE protein from histidine to aspartic acid at position 63, known herein as H63D. Primers H63D_WT_AS and H63D_WT_OP were designed to direct the amplification of a 121 bp product from the normal *HFE* gene, with a calculated T_m of 86.0°C (wild type primer set). H63D_WT_AS has a 'C' nucleotide at the 3' end matching the 'G' base on the non-coding strand of the *HFE* gene. H63D_MUT_AS and H63D_MUT_OP amplify an 89 bp fragment from homozygous mutant DNA, with a calculated T_m of 84.0°C (mutant primer set). H63D_MUT_AS has a 'C' nucleotide at the 3' end, matching the 'G' polymorphic base on the coding strand of the *HFE* gene. The sequence of interest and primer alignments is shown in Figure 4.4.

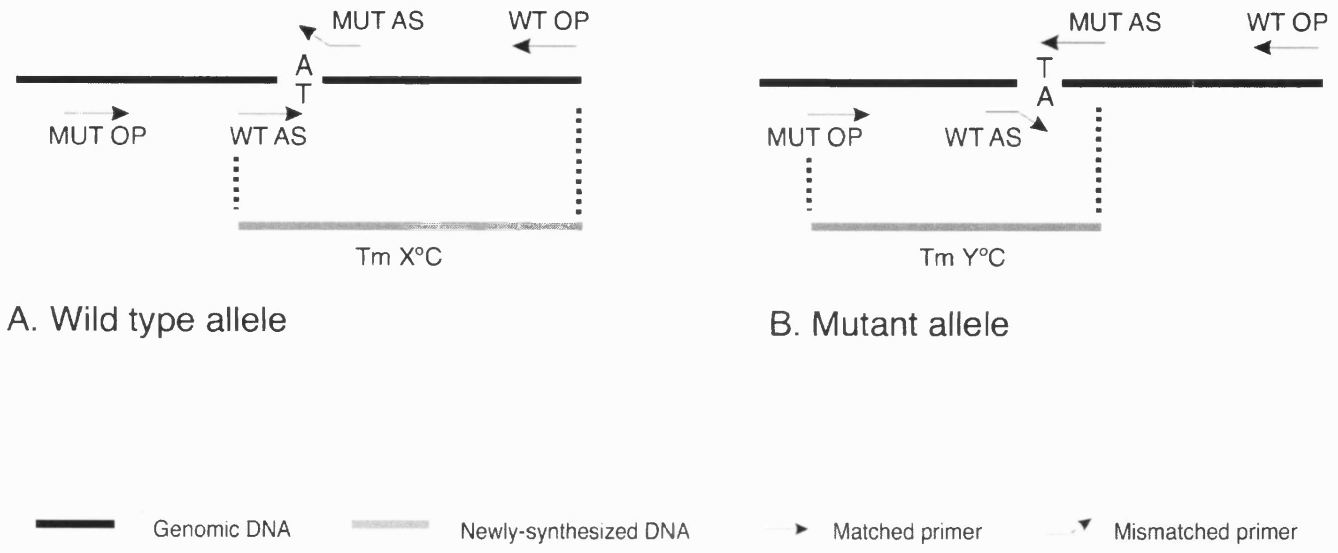


Figure 4.1: A schematic representation of 'fluorescent bi-directional PCR'. In a single PCR reaction with two allele-specific primer sets, genomic DNA will be amplified by either set depending upon template zygosity, yielding allele-specific amplicons. [A] Samples homozygous for the wild type allele will be amplified by WT_AS and WT_OP giving an amplicon with a pre-defined T_m of X°C. [B] Samples homozygous for the mutant allele will be amplified by MUT_AS and MUT_OP generating an amplicon with a pre-defined T_m of Y°C. The difference between X and Y must be 2°C for adequate discrimination using instruments available currently. Heterozygous inheritance will be diagnosed by amplification of both allele-specific PCR products.

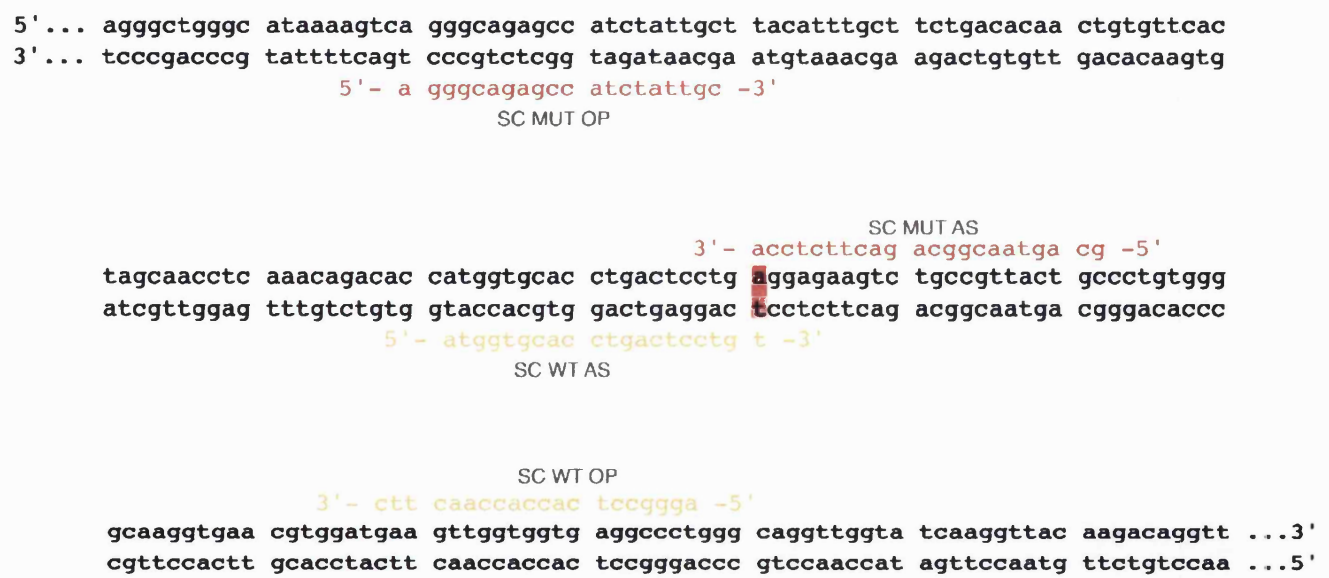


Figure 4.2: Sequence of interest and primer alignments for the detection of sickle cell anaemia (NCBI Accession Number M34404) using the fluorescent bi-directional PCR assay. Area highlighted in red represents the wild type sequence at the SNP site. The mutant allele would have a polymorphic 'T' on the coding strand with a complementary 'A' on the non-coding strand.

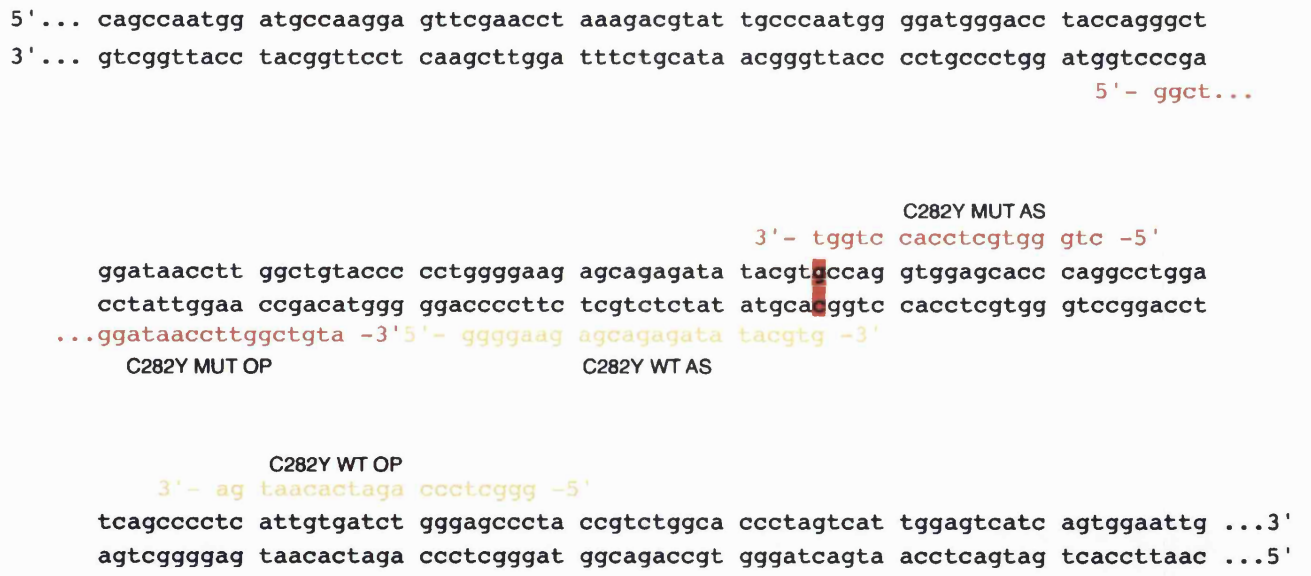


Figure 4.3: Sequence of interest and primer alignments for the detection of the C282Y mutation in the *HFE* gene (NCBI Accession Number Hs. 20019). Area highlighted in red represents the wild type sequence at the SNP site. The mutant allele would have a polymorphic 'A' on the coding strand with a complementary 'T' on the non-coding strand.

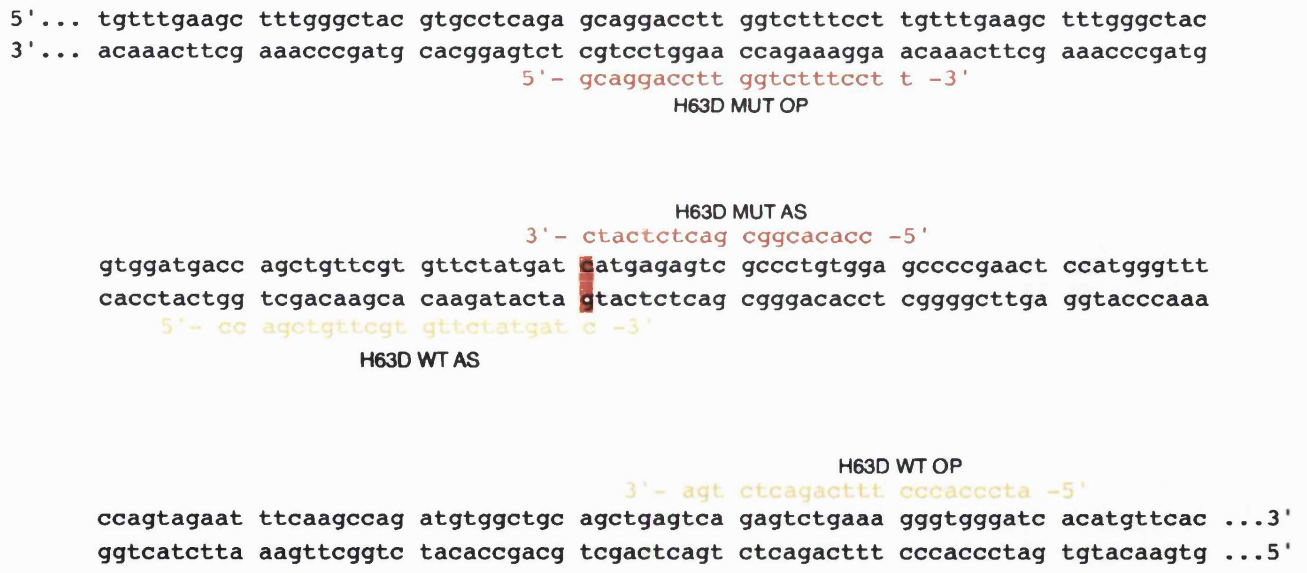


Figure 4.4: Sequence of interest and primer alignments for the detection of the H63D mutation in the *HFE* gene (NCBI Accession Number Hs. 20019). Area highlighted in red represents the wild type sequence at the SNP site. The mutant allele would have a polymorphic 'G' on the coding strand with a complementary 'C' on the non-coding strand.

4.1.2 PCR amplifications

All PCR tests were a total volume of 20 µl, containing 2 µl of LightCycler™ DNA FastStart SYBR Green I mix, oligonucleotides, PCR grade water, a final concentration of 3 mM $MgCl_2$ and 5-20 ng of human genomic DNA. Unknown DNA samples were received as extracted DNA samples (from blood), or were isolated directly from blood (Section 2.5.1) or a buccal cell swab (Section 2.5.2). PCR cycling and melt analysis protocols were performed on the LightCycler™ PCR instrument according to Section 2.0.2. The thermal cycling profile consisted of an initial denaturation step (10 min) at 95 °C, followed by 35 amplification cycles of 1 s denaturation at 95 °C, and 3 s annealing at 62 °C. A single fluorescence reading was acquired as the sample passed through 72°C. For assays designed to detect the C282Y and H63D mutations, the annealing time was reduced to 1 s. Product identification and specificity was confirmed by performing a melting curve analysis according to Section 2.0.3.

4.1.3 Assay Optimisation

The optimisation strategy described in Figure 4.5/ Table 4.2 was employed to define optimal primer ratios for distinguishing between wild type and mutant genotypes. An initial magnesium titration was performed testing a range of concentrations (1, 2, 3, 4 and 5 mM) in PCR tests directed by outer primers only. In each case, a single product without artefactual amplification was yielded with a final magnesium chloride concentration of 3 mM, which was used in subsequent reactions.

Figure 4.5: General optimisation strategy for defining discriminatory conditions for fluorescent bi-directional PCR tests.

Rxn	*WT AS, μM*	*WT OP, μM*	*MUT AS, μM*	*MUT OP, μM*
1	A	A	A	A
2	A	A	B	B
3	A	A	C	C
4	B	B	A	B
5	B	B	B	C
6	B	B	C	A
7	C	C	A	C
8	C	C	B	A
9	C	C	C	B

Table 4.2: Primer titration matrix. Primers were aliquoted according to this matrix, where A= 0.15 μM, B= 0.3 μM and C= 0.9 μM. Inter-set primer concentrations were chosen empirically.

Figure 4.6 displays gel images from optimisation sets incorporating wild type or mutant DNA, used to define optimal discriminatory conditions for the detection of the mutation associated with sickle cell anaemia. Corresponding T_m data is displayed in Table 4.3. Conditions of Reaction 3, Table 4.2 (0.15 μM each SC_WT_AS/ SC_WT_OP, and 0.9 μM each SC_MUT_AS/ SC_MUT_OP) yielded a single product with a T_m of 87.4°C in the presence of wild type DNA, or 85.7°C with mutant DNA. Each melt peak had an area of 17.78 and 26.01 arbitrary fluorescence units, respectively. This corresponded to a single band on the agarose gel at 86 bp for wild type DNA and 112 bp for mutant DNA. These conditions were used in subsequent genotyping tests for detecting the presence of the A→T mutation.

Figure 4.7 shows gel images from optimisation sets incorporating wild type or mutant DNA, used to define optimal discriminatory conditions for the detection of the C282Y polymorphism. Corresponding T_m data is displayed in Table 4.4. Conditions of Reaction 5, Table 4.2 (0.3 μM each C282Y_WT_AS/ C282Y_WT_OP, and 0.3 μM each C282Y_MUT_AS/ C282Y_MUT_OP) yielded a product with a T_m of 84.15°C in the presence of wild type DNA, and 87.81°C with mutant control DNA. This corresponded to a 76 bp and 72 bp fragment respectively. The larger 118 bp fragment amplified by outer primers was also present in Reaction 5 containing wild type DNA, with a melting temperature of 88.58°C.

A second round of optimisation was performed to refine conditions further (Figure 4.8, Table 4.5). In this case, 0.25 μM each primer set was used to genotype six control DNA samples; two wild type, two mutant and two heterozygotes. A single 76 bp product was formed for each wild type homozygote with a T_m of 87.28 and 87.38 °C. A single 72 bp product was formed for each mutant homozygote with a T_m of 84.84 and 84.95 °C. Two products of both pre-defined sizes were formed with the heterozygous control DNA samples without artefactual amplification. These conditions were used for subsequent genotyping tests for detecting the C282Y mutation.

Figure 4.9 shows gel images from optimisation sets incorporating wild type or mutant DNA, used to define optimal discriminatory conditions for the detection of the H63D polymorphism. Corresponding T_m data is displayed in Table 4.6. Conditions of Reaction 5, Table 4.2 (0.6 μM each primer) yielded a product with a T_m of 87.18°C in the presence of wild type DNA, or 85.08 °C with mutant DNA. Each melt peak had an area of 19.33 and 29.12 arbitrary fluorescence units, respectively. This corresponded to a band on the agarose gel at 121 bp for wild type DNA and 89 bp for mutant DNA. A small amount of target region amplification was detected in the wild type sample.

A further round of optimisation was performed using a narrowed range of primer concentrations (Table 4.2, where A= 0.2 μM, B= 0.25 μM and C= 0.3 μM). Due to limiting volumes of control DNA samples, a heterozygous control DNA was used as template. Reaction 7, Table 4.2 (0.3 μM of each H63D_WT_AS/ H63D_WT_OP and 0.2 μM each H63D_MUT_AS/ H63D_MUT_OP) amplified two allele-specific fragments to similar levels (Figure 4.10, Table 4.7). The wild type melt peak (87.19°C) had an area of 16.89 arbitrary fluorescence units, and the mutant melt peak (84.16°C) had an area of 14.38 arbitrary fluorescence units. This corresponded to a band on the agarose gel at 121 bp for wild type DNA and 89 bp for mutant DNA. To confirm the optimal nature of these conditions, all control samples were re-tested. Gel images are displayed in Figure 4.11, with corresponding T_m data in Table 4.8. Known homozygous mutant samples were indistinguishable from heterozygotes, both generating identical band patterns containing both wild type (121 bp) and mutant (89 bp) fragments with similar T_m values (wild type, 87.82 ±0.05°C and mutant, 85.13 ±0.16°C).

To confirm whether the ambiguity was due to the reaction conditions or a possible sample mislabelling, the reference laboratory was contacted to disclose two additional DNA samples that could be used as homozygous controls. These samples were used as template in PCR tests using the conditions of Reaction 7 (Table 4.2), and fragment patterns were compared with those from the original mutant control. The resulting gel image and T_m data is shown in Figure 4.12 and

Table 4.9, respectively. The band profile of the original control differed from the two additional mutant controls. As previous, a wild type amplicon (121 bp, 87.54°C) and a mutant amplicon (89 bp, 84.17°C) was amplified from the original mutant control DNA, whilst single 89 bp amplicons (were amplified with the new control samples (84.70°C and 85.18°C), confirming that the original mutant control sample was in fact heterozygous for the H63D mutation. These conditions were used in subsequent genotyping tests.

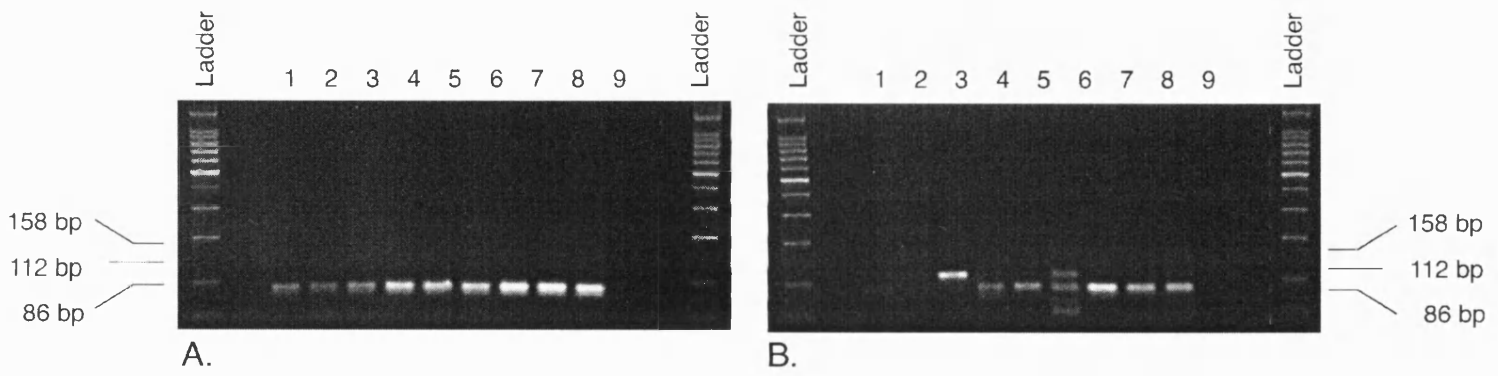


Figure 4.6: Typical gel images of optimisation titrations from the assay designed to detect the presence of the sickle cell mutation. Reactions 1 to 9 correspond to conditions described in Table 4.2 using [A] Wild type and, [B] Mutant human genomic DNA template. The 158 bp target region was not amplified to a detectable level across the reaction set. Corresponding T_m data is displayed in Table 4.3 overleaf.

Gel [A] Wild type

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (158 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wild type (112 bp)	87.25	16.86	87.48	15.69	87.36	17.78	87.06	24.57	87.19	27.82	87.21	27.64	87.16	31.76	87.17	32.94	87.13	29.41
Mutant (86 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Gel [B] Mutant

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (158 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wild type (112 bp)	87.45	2.94	87.13	5.63	-	-	87.20	15.66	86.79	15.98	87.23	15.57	86.52	26.79	86.64	21.96	86.85	26.09
Mutant (86 bp)	-	-	-	-	85.66	26.01	-	-	-	-	85.35	4.47	-	-	-	-	-	-

Table 4.3: T_m data from optimisation sets used to detect the presence of the sickle cell mutation, using [A] wild type or [B] mutant DNA template. Corresponding gel images are shown in Figure 4.6.

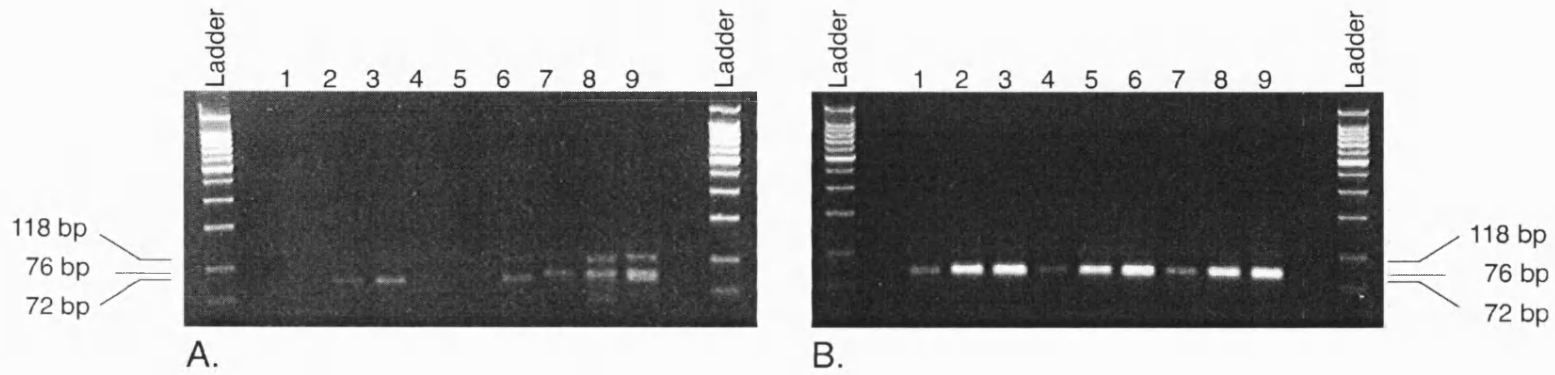


Figure 4.7: Typical gel images of optimisation titrations from the assay designed to detect the presence of the C282Y mutation. Reactions 1 to 9 correspond to conditions described in Table 4.2 using [A] Wild type and, [B] Mutant human genomic DNA template. The 118 bp target region was amplified to a detectable level in Reactions 8 and 9 of Set [A]. Corresponding T_m data is displayed in Table 4.4 overleaf.

Gel [A] Wild type

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (118 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	88.76	5.86	88.18	22.20
Wild type (76 bp)	87.97	1.19	-	-	-	-	87.19	15.91	87.81	20.72	88.6	16.00	87.03	22.23	86.95	20.93	-	-
Mutant (72 bp)	-	-	84.45	15.39	84.13	23.81	-	-	-	-	83.52	15.73	-	-	-	-	-	-

Gel [B] Mutant

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (118 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wild type (76 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mutant (72 bp)	84.65	13.82	84.40	29.90	84.44	33.32	84.43	8.62	84.15	22.47	84.43	33.34	84.09	9.30	84.06	21.03	84.26	28.90

Table 4.4: T_m data from optimisation sets used to detect the presence of the C282Y mutation using [A] wild type or [B] mutant DNA template. Corresponding gel images are shown in Figure 4.7.

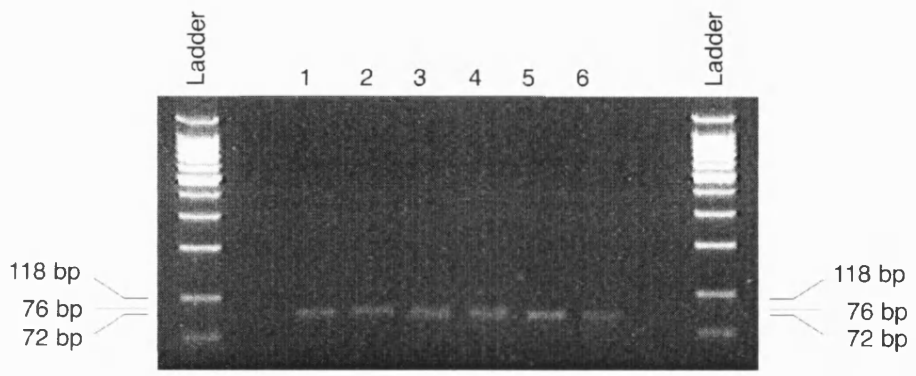


Figure 4.8: Typical gel images from further titrations for optimisation of the assay designed to detect the C282Y mutation. Lanes 1 and 2 contain wild type DNA samples; Lanes 3 and 4 contain heterozygous DNA samples; Lanes 5 and 6 contain mutant DNA samples. All DNA samples used were different. The 118 bp target region was not amplified to a detectable level across the reaction set. Corresponding T_m data is displayed in Table 4.5 overleaf.

Reaction	*1*		*2*		*3*		*4*		*5*		*6*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (118 bp)	-	-	-	-	-	-	-	-	-	-	-	-
Wild type (76 bp)	87.28	2.17	87.38	2.50	87.43	1.41	87.54	1.62	-	-	-	-
Mutant (72 bp)	-	-	-	-	84.67	2.58	84.72	3.02	84.84	4.76	84.95	2.55

Table 4.5: T_m data from a second optimisation set used to detect the presence of the C282Y mutation. Corresponding gel images are shown in Figure 4.8.

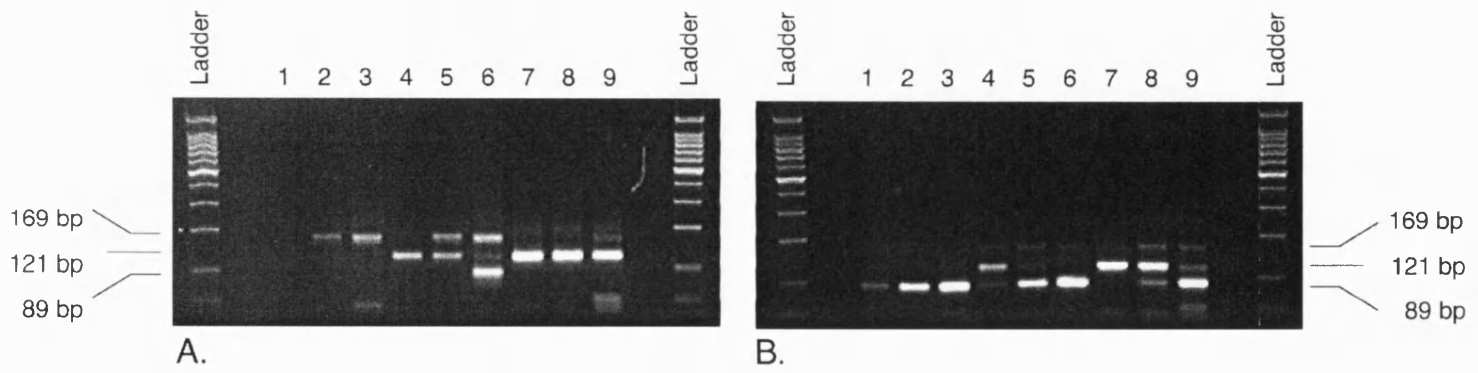


Figure 4.9: Typical gel images of optimisation titrations from the assay designed to detect the presence of the H63D mutation. Reactions 1 to 9 correspond to conditions described in Table 4.2 using [A] Wild type and, [B] Mutant human genomic DNA template. The 169 bp target region was amplified to a detectable level in Reactions 2, 3, 5 and 6 of Set [A], and Reactions 5, 8 and 9 of Set [B]. Corresponding T_m data is displayed in Table 4.6 overleaf.

Gel [A] Wild type

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (169 bp)	87.24	1.06	87.82	6.86	87.60	15.81	-	-	IND	-	87.63	-	-	-	-	-	-	-
Wild type (121 bp)	-	-	-	-	-	-	86.96	17.62	87.18	19.33	-	19.25	86.57	31.86	86.48	31.88	86.67	32.05
Mutant (89 bp)	-	-	-	-	-	-	-	-	-	-	IND	-	-	-	-	-	-	-

Gel [B] Mutant

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (169 bp)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wild type (121 bp)	-	-	-	-	-	-	87.65	16.53	-	-	-	-	87.29	31.51	87.36	33.55	88.09	8.85
Mutant (89 bp)	85.68	10.32	85.12	37.11	85.10	44.63	-	-	85.08	29.12	85.15	43.24	-	-	-	-	84.78	31.36

Table 4.6: T_m data from optimisation sets used to detect the presence of the H63D mutation using [A] wild type or [B] mutant DNA template. Corresponding gel images are shown in Figure 4.9.

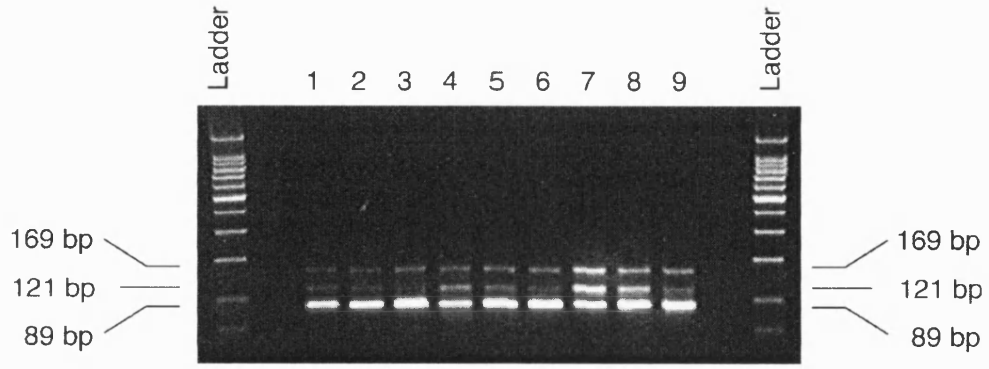


Figure 4.10: A typical gel image of an optimisation titration from the assay designed to detect the presence of the H63D mutation. Reactions 1 to 9 correspond to conditions described in Table 4.2 using heterozygous human genomic DNA template. The 169 bp target region was amplified to a detectable level in all reactions. Corresponding T_m data is displayed in Table 4.7 overleaf.

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (169 bp)	-	-	-	-	IND	-	IND	-	IND	-	IND	-	IND	-	IND	-	IND	-
Wild type (121 bp)	87.57	4.75	87.72	3.10	87.83	3.96	87.37	9.05	87.63	7.07	87.78	4.49	87.19	16.89	87.46	11.50	87.64	7.59
Mutant (89 bp)	84.58	22.91	84.57	27.43	84.46	30.11	84.44	17.13	84.46	25.06	84.44	28.08	84.16	14.38	84.37	21.91	84.39	28.26

Table 4.7: T_m data from a second optimisation set used to detect the presence of the H63D mutation. Corresponding gel images are shown in Figure 4.10.

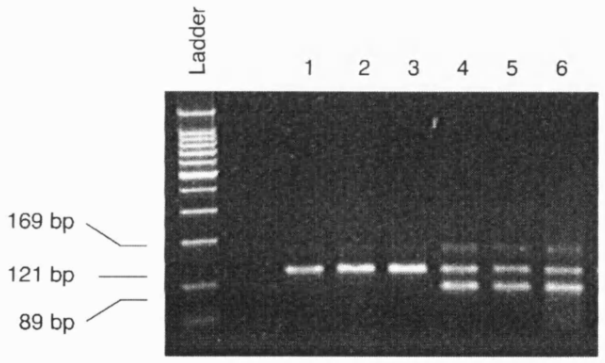


Figure 4.11: A typical gel image using optimised conditions for the detection the H63D mutation. Lanes 1, 2 and 3 contain wild type DNA samples; Lanes 4 and 5 contain heterozygous DNA samples; and lane 6 contains mutant DNA samples. All DNA samples used were different. The 169 bp target region was amplified to a detectable level in Reactions 4, 5 and 6. Corresponding T_m data is displayed in Table 4.8 overleaf.

Reaction	*1*		*2*		*3*		*4*		*5*		*6*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (169 bp)	-	-	-	-	-	-	IND	-	IND	-	IND	-
Wild type (121 bp)	87.54	12.76	87.51	14.79	87.48	17.80	87.76	10.83	87.86	7.75	87.83	10.45
Mutant (89 bp)	-	-	-	-	-	-	85.07	12.65	85.32	10.26	85.01	14.54

Table 4.8: T_m data from control sets used to clarify the conditions established for the detection of the H63D mutation. Corresponding gel images are shown in Figure 4.11.

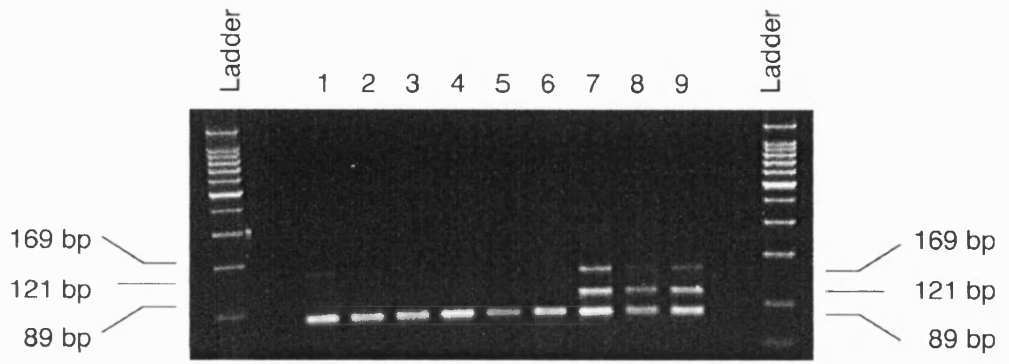


Figure 4.12: Typical gel images using optimised conditions for the detection the H63D mutation. Lanes 1-3 contains a mutant DNA sample; lanes 4-6 contain another mutant DNA sample; and lanes 7-9 contain the control mutant template used in previous optimisations of this assay. This sample has been revealed here as a heterozygote. Lanes 1, 4 and 7 contained 20 ng starting DNA concentration; lanes 2, 5 and 8 contained 10ng; and lanes 3, 6, and 9 contained 5 ng. Less target region amplification was observed in the heterozygote sample where a lower starting DNA concentration was included. This will be accommodated in genotyping experiments. Corresponding T_m data is displayed in Table 4.9 overleaf.

Reaction	*1*		*2*		*3*		*4*		*5*		*6*		*7*		*8*		*9*	
Amplicon	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area	T_m, °C	Area
Target region (169 bp)	-	-	-	-	-	-	-	-	-	-	-	-	IND	-	IND	-	IND	-
Wild type (121 bp)	-	-	-	-	-	-	-	-	-	-	-	-	87.54	19.57	88.07	8.42	87.89	13.71
Mutant (89 bp)	84.70	28.35	85.31	23.06	85.51	21.21	85.18	26.37	85.87	13.55	85.57	19.94	84.17	15.95	85.26	14.02	84.76	17.36

Table 4.9: T_m data from control sets used to clarify the conditions established for the detection of the H63D mutation. Corresponding gel images are shown in Figure 4.12.

4.1.4 SNP genotyping

For the sickle cell assay, forty-two DNA samples were tested in a small-blinded trial. Thirty-three samples generated a melt peak around 87.5°C (±0.2°C), five samples produced a melt peak at 85.7°C (±0.2°C), and seven samples amplified both allele-specific fragments giving a dual-peak with maxima at 85.3°C (±0.3°C) and 87.6°C (±0.3°C); thus diagnosing homozygous wild type, homozygous mutant and heterozygous inheritance respectively (Figure 4.13). Cluster analysis of product T_m values showed three unambiguous and distinct clustering regions, each corresponding to a particular genotype (Figure 4.14).

For the C282Y assay, a total of sixty-two DNA samples were genotyped. Thirty-two samples generated a melt peak around 87.2°C (±0.3°C), seventeen samples produced a melt peak at 84.6°C (±0.3C), and thirteen samples amplified both allele-specific fragments giving a dual-peak with maxima at 84.5°C (±0.3°C) and 87.4°C (±0.2°C); thus diagnosing homozygous wild type, homozygous mutant and heterozygous inheritance respectively (Figure 4.15). Cluster analysis of product T_m values showed three unambiguous and distinct clustering regions, each corresponding to a particular genotype (Figure 4.16).

For the H63D assay, sixty-two DNA samples were genotyped. Thirty-nine samples generated a melt peak around 87.7°C (±0.3°C), ten samples produced a melt peak at 85.3°C (±0.1°C), and thirteen samples amplified both allele-specific fragments giving a dual-peak with maxima at 85.1°C (±0.4°C) and 87.7°C (±0.3°C); thus diagnosing homozygous wild type, homozygous mutant and heterozygous inheritance respectively (Figure 4.17). Cluster analysis of product T_m values showed three unambiguous and distinct clustering regions, each corresponding to a particular genotype (Figure 4.18). For each test, 'no template control' samples were included to assess contamination and non-specific amplification by primer sets; no amplification occurred in these controls. If primer dimer had been formed in the reaction, the characteristically low T_m of the artefact would ensure that points were clearly separate from profiles generated by actual template amplifications.

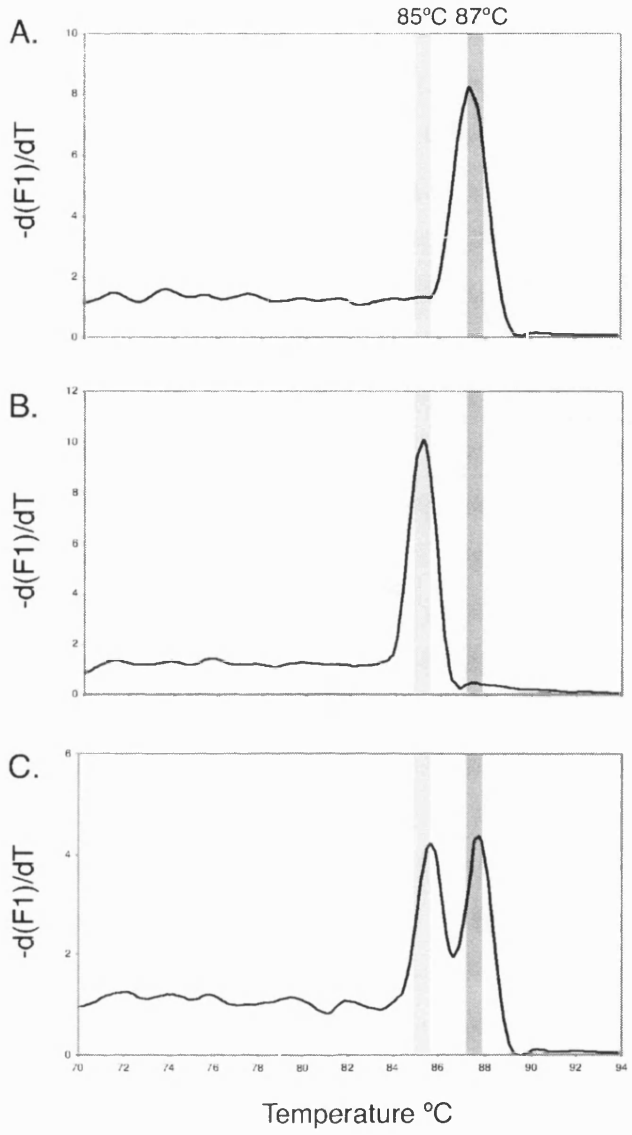


Figure 4.13: Example melting profiles from diagnostic tests for the detection of the sickle cell mutation representing typical results from one of each possible genotype; [A] Homozygous wild type, [B] Homozygous mutant, and [C] Heterozygous DNA. The three genotypes are easily identified by visual inspection of the graphs as described in text. Actual T_m values were slightly higher than calculated in most cases.

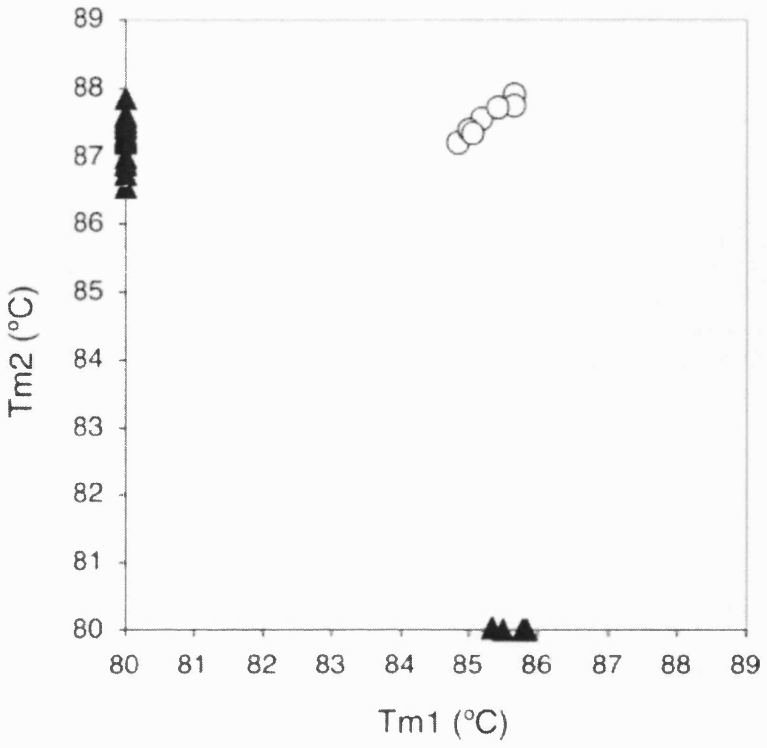


Figure 4.14: Cluster analysis to determine sickle cell genotype. The x-axis represents peaks within the actual T_m range of 85.4°C (±0.3°C), and the y-axis represents areas within the actual T_m range of 87.3°C (±0.3°C). A symbol is assigned to represent homozygous DNA (▲) or heterozygous DNA (O). Points lining up along either the x-axis or y-axis diagnose homozygous inheritance of mutant or wild type alleles, respectively. Points lying near to the x=y axes represents heterozygous inheritance.

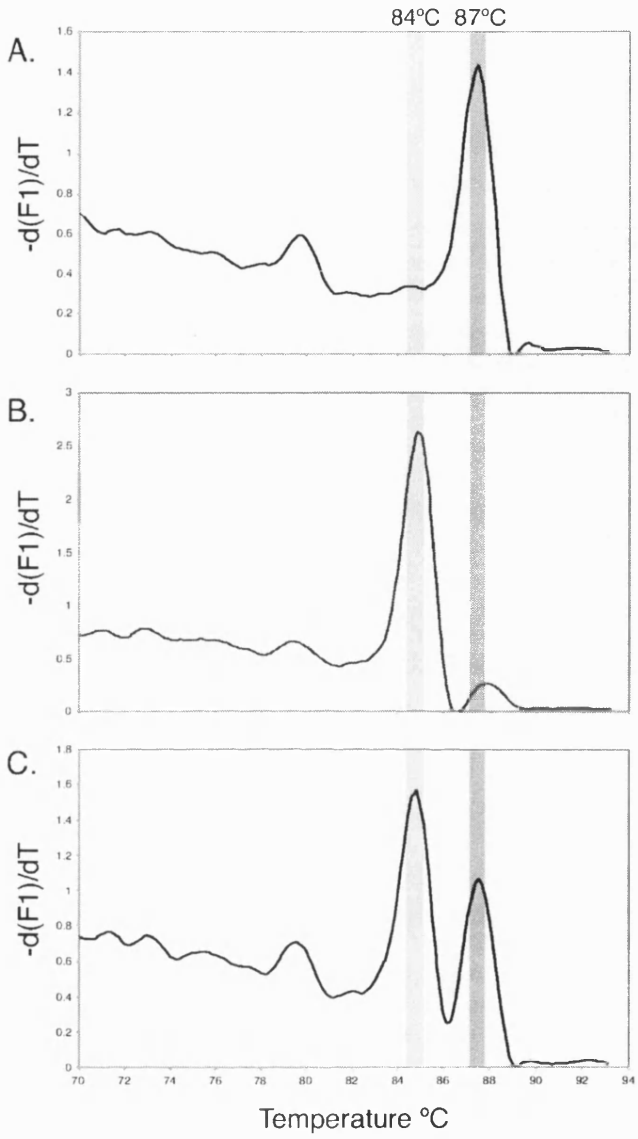


Figure 4.15: Example melting profiles from diagnostic tests for the detection of the C282Y mutation representing typical results from one of each possible genotype; [A] Homozygous wild type, [B] Homozygous mutant, and [C] Heterozygous DNA. The three genotypes are easily identified by visual inspection of the graphs as described in text. Actual T_m values were slightly higher than calculated in most cases.

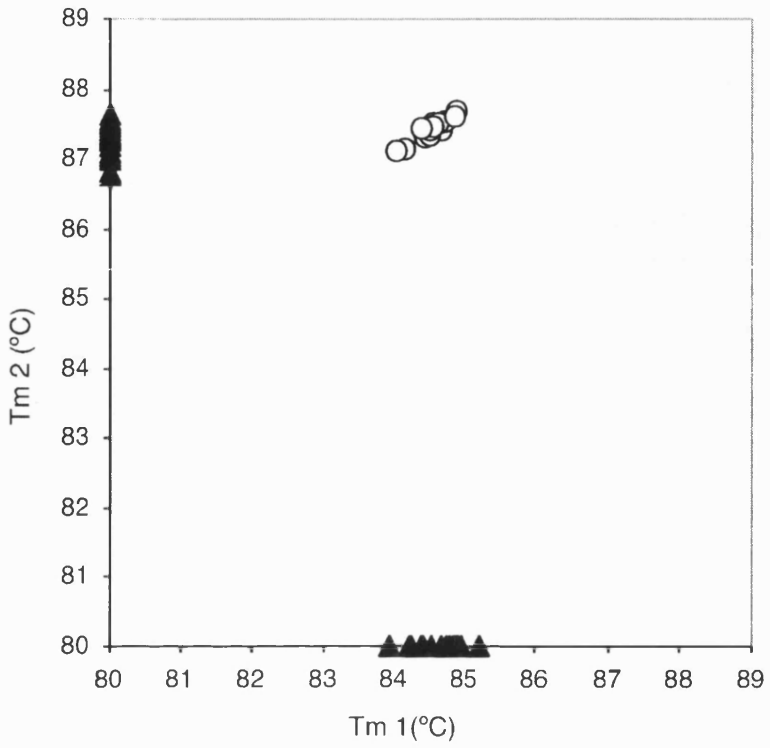


Figure 4.16: Cluster analysis to determine C282Y HFE genotype. The x-axis represents peaks within the actual T_m range of 87.2 °C (± 0.3°C), and the y-axis represents areas within the actual T_m range of 84.6 °C (± 0.3°C). A symbol is assigned to represent homozygous DNA (▲) or heterozygous DNA (O). Points lining up along either the x-axis or y-axis diagnose homozygous inheritance of mutant or wild type alleles, respectively. Points lying near to the x=y axes represents heterozygous inheritance.

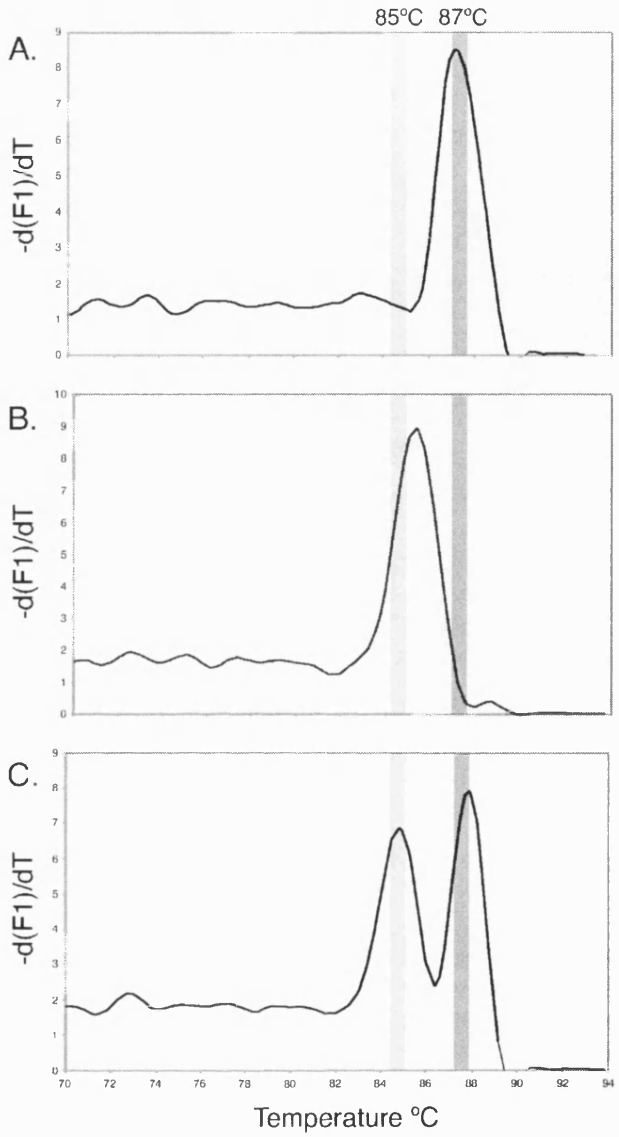


Figure 4.17: Example melting profiles from diagnostic tests for the detection of the H63D mutation representing typical results from one of each possible genotype; [A] Homozygous wild type, [B] Homozygous mutant, and [C] Heterozygous DNA. The three genotypes are easily identified by visual inspection of the graphs as described in text. Actual T_m values were slightly higher than calculated in most cases.

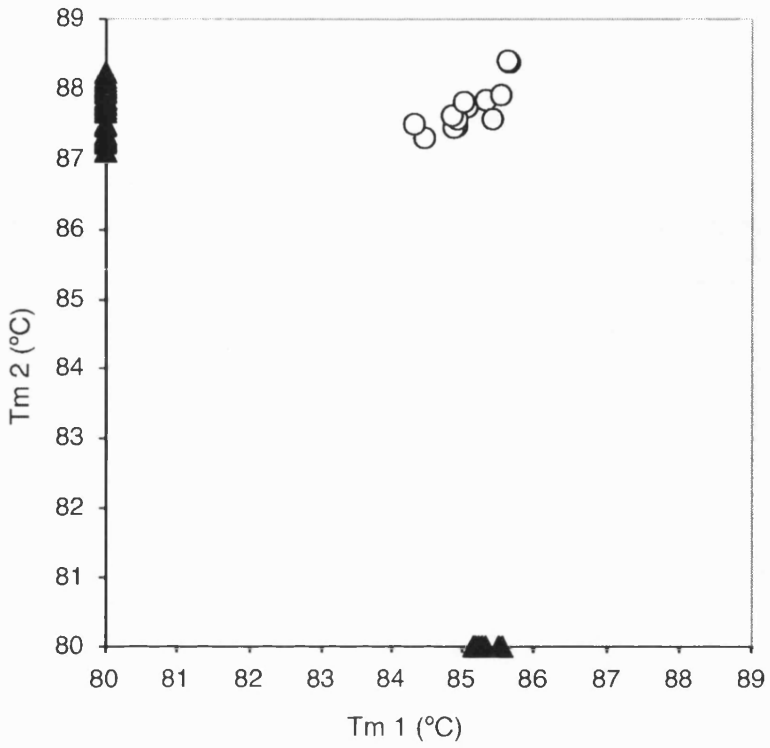


Figure 4.18: Cluster analysis to determine H63D HFE genotype. The x-axis represents peaks within the actual T_m range of 87.7 °C (± 0.3°C), and the y-axis represents areas within the actual T_m range of 85.3°C (± 0.1°C). A symbol is assigned to represent homozygous DNA (▲) or heterozygous DNA (O). Points lining up along either the x-axis or y-axis diagnose homozygous inheritance of mutant or wild type alleles, respectively. Points lying near to the x=y axes represents heterozygous inheritance.

4.1.5 Genotype validation

For the sickle cell genotyping trial, SNP scoring by fluorescent bi-directional PCR was validated by restriction enzyme digestion according to the protocol described in Section 2.5.4. PCR products directed by outer primers (SC_WT_OP and SC_MUT_OP) were cut with *Dde*I, which has a specific cutting site of 5'- CTNAG -3' (where N is G, C, A, or T; Chang and Kan, 1982). The presence of the sickle cell mutation abolishes the cutting site for this enzyme. A large uncut fragment was obtained for mutant samples, as opposed to two smaller fragments for a normal genotype. The presence of all three fragments diagnosed a heterozygous genotype. RFLP results were 100% concordant with the fluorescent bi-directional PCR assay (Figure 4.19).

For HH assays, the fluorescent bi-directional PCR genotyping trial was completed prior to receipt of actual genotypes from the source reference laboratory. The reference genotyping method used was a multiplex PCR followed by SSCP on a polyacrylamide gel at 8°C, visualised by silver staining (Unpublished method, Steve Keeney, Manchester Royal Infirmary, UK). Genotypes called by fluorescent bi-directional PCR were compared for concordance. Data from the C282Y assay were 92% concordant with genotyping data from the reference laboratory. Results from the H63D assay were 97% concordant. Samples giving different results were sequenced at Bath University using Big Dye Terminator Cycle Sequencing Ready Reaction Kits on an ABI Prism 377 DNA Sequencer (Applied Biosystems). In each case, the sequence data corresponded to the fluorescent bi-directional assay.

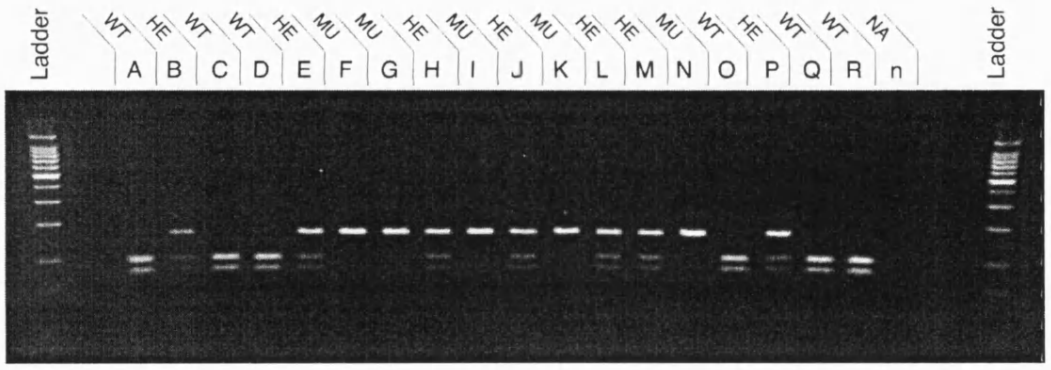


Figure 4.19: Genotype validation by restriction enzyme digestion. Samples A to R are displayed with a no template control sample (n).

4.2 Discussion

The fluorescent bi-directional PCR assay described here was successfully used to genotype human genomic DNA samples for the presence of three well-characterised single nucleotide polymorphisms. The integration of the single tube bi-directional PCR (Chapter three), with SYBR Green I, melt curve analysis and rapid cycling has enabled favourable interactions addressing key issues facing developers of clinical diagnostic assays, including specificity, throughput, and cost.

The design of a bi-directional PCR test is quite simple. The resolution limits of current instrumentation demands that in order for two products to be accurately separated during melt curve analysis using a dsDNA specific binding dye, their melting temperatures must differ by at least 2°C (Ririe *et al.,* 1997). Therefore, the test requires the design of primer sets that satisfy this parameter. Generally, this can be achieved by designing smaller amplicons. Empirical formulas predict that a DNA duplex made up entirely of A and T nucleotides would melt 41°C lower than a 100% GC duplex. Given the same GC content, a 40-base-pair primer dimer should melt 12°C below a 1000-bp product. Hence, the theoretical range for potential PCR products spans at least 50°C (Ririe *et al.,* 1997). Using the β-globin gene sequence as an example, Figure 4.20 shows that in practice, this range only holds for smaller amplicons between 50-150 bp. Providentially, this is the range of choice for rapid cycling protocols due to the short extension times (Wittwer *et al.,* 1993).

For the detection of each mutation, the GC content of the sequence surrounding the SNP site enabled the design of primers that amplified products with a predicted T_m difference of 2°C (Figures 4.2-4.4 inclusive). Detection of the sickle cell SNP was aided by the high GC content downstream of the SNP site that formed the wild type fragment (58%) compared to the upstream sequence that formed the mutant fragment (49%). Likewise, primers designed for the detection of HH mutations, was based on the differing GC contents of the target regions; this was 56% upstream and 63% downstream of the C282Y SNP site; and 49% upstream and 53% downstream of the H63D site.

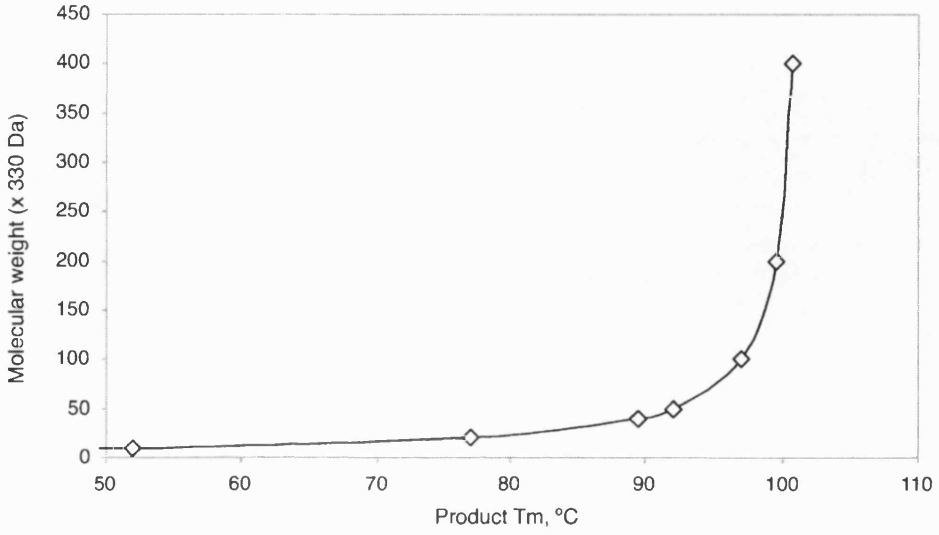


Figure 4.20: Relationship between molecular weight and product T_m of sequence surrounding sickle cell SNP in ß-globin gene, as an example. This graph was constructed by sequentially measuring the T_m of different sized fragments within the β-globin gene, using PrimerCalc Software Version 1.0; Q-Biogene), which employs nearest neighbour thermodynamic algorithms (Breslauer *et al.*, 1986).

The relationship between amplicon size and product T_m has consequence in other aspects of this assay. During bi-directional PCR, *Taq* DNA polymerase will extend the two non allele-specific primers flanking the target region regardless of genotype, generating a 158, 103, or 169 bp amplicon with a calculated T_m of ≈ 88.0°C, 87.0°C, and 88.0°C from sickle cell and *HFE* genes (C282Y and H63D), respectively. However, the melt peak from this larger fragment is indistinguishable from the largest allele-specific amplicon, because there can only be a small difference in their T_m. Amplification of this fragment may lead to false results by masking allele-specific amplifications. By performing a two-step PCR with denaturation and annealing stages only, cycling conditions were adjusted to favour the amplification of smaller amplicons, relying on the transitional times of the kinetic PCR cycle (De Silva and Wittwer, 2000). This rapid protocol was effective in suppressing amplification of the larger target region in all assays (Figure 4.6, 4.7-4.12 inclusive).

Rapid cycling protocols confer further advantages to AS-PCR methods by improving specificity (De Silva *et al.,* 1998). Traditionally, allele-specific PCR tests have been designed for Peltier-based PCR instruments requiring ~30 s extension times. This increases the time for possible extension of a mismatched primer annealed at the expected site to occur. If the conditions have not been effectively optimised, an amplicon of identical size to that formed by a perfectly matched primer may be produced, causing failure of the genotyping test. Using air-exchange as the heating and cooling mechanism, faster ramp rates and improved temperature control can be achieved, allowing shorter extension times which have been shown to strongly disfavour mismatched primer extension (De Silva *et al.,* 1998; Wittwer *et al.,* 1993). This is likely due to the decreased time available for a mismatched primer to 'sit' at the annealing site, reducing the chance of extending through. Consequently, fluorescent bi-directional PCR works effectively without primer modifications such as GC tails or extra mismatches near the 3' end of the primers to improve specificity (Germer and Higuchi, 1999).

A simple optimisation strategy was used to define the conditions taking into account the assay dynamics described above (Figure 4.5). Since the commercial mastermix used contained the majority of reaction components, optimisation involved a simple

sequential titration of Mg^{2+} and oligonucleotide. Discriminatory conditions were specified as those able to amplify allele-specific products *only* with similar efficiencies determined by melt peak height and peak area. In addition, these conditions should reduce or completely remove artefactual amplification (e.g. primer dimer). Ideally, only two oligonucleotide titration sets using wild type or mutant control DNA samples would be required to define optimal primer concentrations for allelic discrimination, and this was the case for the assay designed to detect the SNP causing sickle cell anaemia. Two sequential rounds of optimisation using a narrower range of primer concentrations or shortened cycling programs were necessary for both HH assays. With the use of novel chemistries such as Locked Nucleic Acids (Jensen *et al.,* 2001), with the ability to improve discrimination in AS-PCR (as shown in Appendix V data) robustness may be further enhanced, eliminating the need for sequential optimisation steps. Since the true cost of genotyping tests is not only defined by reagent and consumable costs, but also the time and labour involved in reaction design and optimisation, the strategy used here added to overall cost effectiveness.

The power of the optimisation strategy was demonstrated by the assay designed to detect the H63D mutation. In this case, mislabelling of a control sample at source meant that conditions were being optimised for wild type and heterozygous samples only (Figure 4.11). The conditions selected were able to identify that the mutant control was in fact heterozygous. This was confirmed by contacting the source laboratory to identify two more mutant samples to which the control could be compared, confirming that it had been correctly genotyped by fluorescent bi-directional PCR (Figure 4.12).

The mastermix used increased the specificity of allele-specific PCR since an aptamer-based hot start enzyme system was incorporated (Table 3.5, pg. 90), preventing non-specific primer extension by *Taq* DNA polymerase at ambient temperatures (Lin *et al.,* 1997). In addition, the use of a commercial mastermix reduced the common observation of well-to-well variation in the T_m of the same amplicon. Such variation is usually attributed to SYBR Green I concentration (Nath *et al.,* 2000), salt concentration (Schildkraut and Lifson, 1965), actual product yield

or temperature transition rates (Ririe *et al.,* 1997). By using a commercial master mix with a consistent concentration of SYBR Green I, and employing optimised magnesium concentrations, this variation was minimised. In practice, the small differences observed were likely due to differences in product yield and smoothing mathematical algorithms and were not sufficient to cause error in fluorescent bi-directional PCR.

In all cases, optimal conditions were defined and validated by genotyping unknown DNA samples for sickle cell anaemia or hereditary haemochromatosis status on the basis of their melting profiles alone. In each case, clearly distinct profiles were achieved depending upon genotype (Figures 4.13, 4.15 and 4.17), which formed distinct clustering regions that may be compatible with SNP-calling software (Figure 4.14, 4.16, 4.18). Excellent concordance levels were achieved with RFLP results in the detection of sickle cell anaemia (Figure 4.19). In the case of HH assays, good concordance was achieved (92% and 97%). The mistyped samples were attributed to sample mix-ups or typographical errors.

To summarise, fluorescent bi-directional PCR is mediated through a highly efficient kinetic process, coupled with an accurate melt curve analysis. The rapid cycling satisfies the higher throughput demanded by today's genotyping needs, and the introduction of homogeneity to the assay is ideal for use for diagnostic testing. The use of a universal dsDNA binding dye maintains the low cost of fluorescent-based assays. The data presented here suggests that fluorescent bi-directional PCR has potential as a tool for screening programmes of sickle cell anaemia or hereditary haemochromatosis, particularly if coupled with informatics tools that would allow automation. With finer temperature control, slower transition rates should be possible as well as greater sample temperature homogeneity (Ririe *et al.,* 1997). This would allow greater resolution, reducing the restrictions for primer design and multiplexing, which represent the only drawbacks of the assay in its current form.

This work was published in BioTechniques, 2002 (Appendix IV).

Chapter Five

Bi-directional PCR Combined with 5' Nuclease Fluorogenic Probes for Allelic Discrimination

5.0 Introduction

The attributes that would lead to a wider acceptance of PCR within clinical diagnostic applications were introduced in detail in Chapter One; the test should be relatively inexpensive, easy and quick to perform, with a high degree of fidelity to limit the likelihood of misdiagnosis (Figure 1.4, pg. 16). Real-time PCR probe technologies go someway towards fulfilling these criteria by allowing sequence specific detection of PCR amplicons as they are synthesised. They also have multiplex capabilities that enable more than one SNP to be assayed within a single reaction, a useful attribute where the interactions of more than one SNP contributes to a clinical phenotype (Taylor *et al.,* 2001), or where a limited source of DNA is available. Homogenous multiplex reactions are able to eliminate the post-PCR processing steps, which are not only rate limiting but also increase the opportunity for sample mix-ups and contamination.

The success of real-time PCR probes is based on hybridisation of a short nucleic acid sequence to its complement; a highly specific molecular recognition event that adds sequence specificity to genotyping assays. Many probe configurations have been described (Figure 1.12, pg. 34), the majority of which utilise FRET as the reporting mechanism. As described in Chapter One, FRET is a distance-dependent interaction between the electronic excited states of two dye molecules in which excitation is transferred from a donor molecule to an acceptor molecule (or quencher) without emission of a photon (Figure 1.6D, pg. 20). The excitation energy is transferred to a neighbouring fluorophore, which raises the electron in the acceptor to a higher energy state as the photo-excited electron in the donor returns to ground state. If the acceptor is also fluorescent, the transferred energy can be emitted as a fluorescence characteristic of the acceptor. If the acceptor is not

fluorescent, the energy is 'quenched' through equilibration with solvent (Didenko, 2000).

The most popular probe configuration utilising FRET for real-time PCR monitoring is the 5' fluorogenic nuclease probe (commercially known as TaqMan® probe; Figure 5.1). A TaqMan® probe consisting of an oligonucleotide specific to sequence between the primer annealing sites is included in the reaction. The probe is labelled with a reporter dye (usually FAM) and a quencher dye (usually TAMRA; 6-carboxy-N,N,N',N'-tetramethylrhodamine), protected from extension by *Taq* DNA polymerase by phosphorylation at the 3' end (Livak, 1999). In the intact probe, the quencher is within close enough proximity to the reporter dye to reduce the fluorescence signal observed by FRET. During the extension phase of a typical PCR run, the 5' nuclease activity of *Taq* degrades the probe in the 5' to 3' direction, liberating the reporter dye from the quencher dye, disabling FRET by increasing the intermolecular distance beyond effective FRET limits, known as the Förster distance. With each additional cycle of denaturation, annealing and extension, a single molecule of the reporter dye is liberated from the quencher dye per PCR product, giving a molar increase in fluorescence signal proportional to increasing amplicon levels.

TaqMan® probes represent a category called 'hydrolysis' probes, named as such because the proximity of the reporter and acceptor dyes is changed by a hydrolysis event. The importance of specific hybridisation using TaqMan® probes is apparent during SNP detection where cleavage of the probe only occurs if it is specifically hybridised to a matched target sequence (Figure 5.2; Livak, 1999). Conventionally, two probes specific for each allele of a bi-allelic SNP are included in the same PCR test, distinguished by labelling with different reporter dyes at the 5' end, emitting fluorescence at different detectable wavelengths. A matched template: probe duplex provides the optimal substrate for 5' nuclease activity, which is a forked structure occurring at the phosphodiester bond joining the displaced region with the base-paired portion of the strand (Figure 1.13, pg. 38; Holland *et al.,* 1991). A single mismatch greatly reduces hybridisation and cleavage efficiency by disruption of this specific recognition structure.

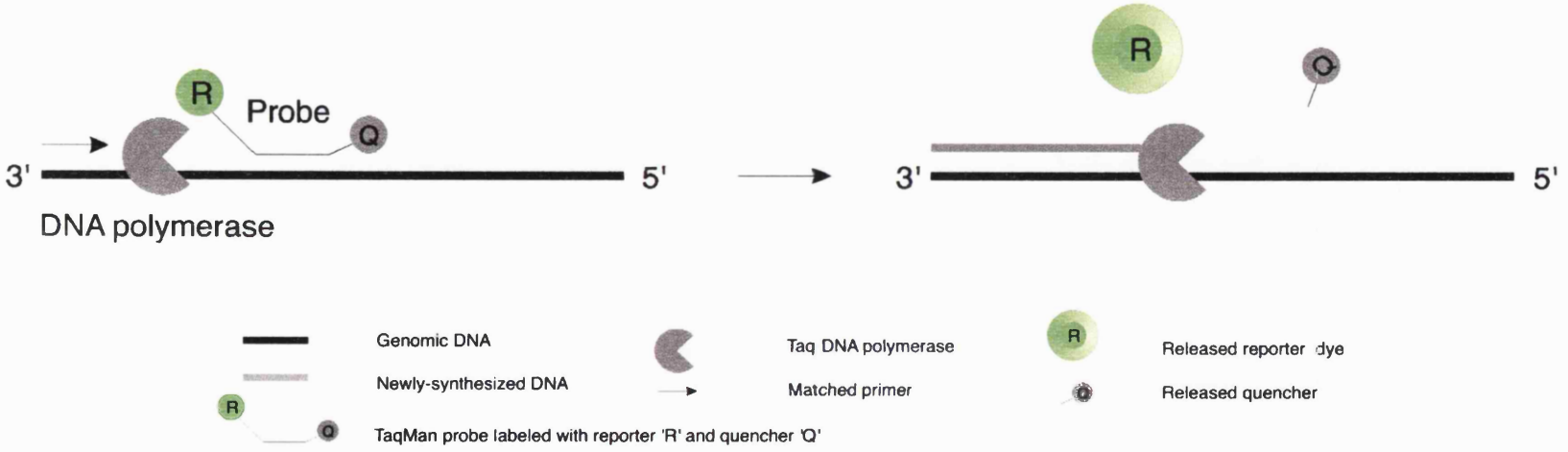


Figure 5.1: Diagram of 5' to 3' exonuclease cleavage of labelled oligonucleotide (TaqMan®) probe. Annealing of the TaqMan® probe to a sequence within the PCR amplicon generates a suitable substrate for 5'→3' exonuclease activity of Taq DNA polymerase. During amplification, the 5'→3' exonuclease activity degrades the probe into smaller fragments, releasing the quencher from the fluorophores. The distance between probe and quencher is no longer within the limits for FRET to occur, and so an increased fluorescent signal from the reporter dye is measured. The signal is proportional to target DNA concentration (Holland *et al.*, 1991).

For a matched TaqMan® assay, fluorescence begins to increase over background levels during early cycles. As described previously, the fractional cycle number at which the fluorescence passes this fixed threshold is known as the "cycle threshold" (C_T; Figure 1.9, pg. 26), indicating that the probe-specific target has been amplified. For genotyping, results are expressed as a signal ratio that reflects the differential hybridisation of the two oligonucleotides to the target sequence (ΔRn); therefore differences in amplification efficiency between samples do not affect interpretation of the genotyping results. An increase in signal from either probe will indicate the dominant alleles in the reaction (Figure 5.3A; Livak, 1999).

For success in conventional assays, a single base mismatch between probe and template has to be detected using a probe that is typically 20-30 nucleotides long. The reported efficacy of TaqMan® here is attributed to a number of factors including a thermodynamic contribution caused by the disruptive effect of a mismatch on hybridisation such that a mismatched probe will have a slightly lower melting temperature than a perfectly matched probe; proper choice of an annealing or extension temperature in the PCR can favour hybridisation of the exact match over a mismatched probe; and the assay is performed under competitive conditions given that both probes are present in the same reaction tube. The stable binding of a matched probe can block hybridisation of the mismatched probe (Livak, 1999).

However, amplification conditions that distinguish between probes with a single nucleotide difference are often difficult to establish (Sevall, 2000). A necessary requirement is that the TaqMan® probe must be bound to the template at temperatures where *Taq* DNA polymerase is optimally extending primers (60-72°C). Probes with melting temperatures between 65°C and 75°C are best. This usually requires an oligonucleotide length of 25-40 nucleotides in order to form the required stable hybrid. This is generally too long for most SNP detections given that the difference in T_m between a perfectly matched duplex 25 bp long and a similar duplex with a single mismatch is usually only about 3°C. Without careful optimisation of probe sequence and PCR conditions it is difficult to detect SNPs with an acceptable signal-to-noise ratio for a robust clinical assay (Walburger *et al.*, 2001).

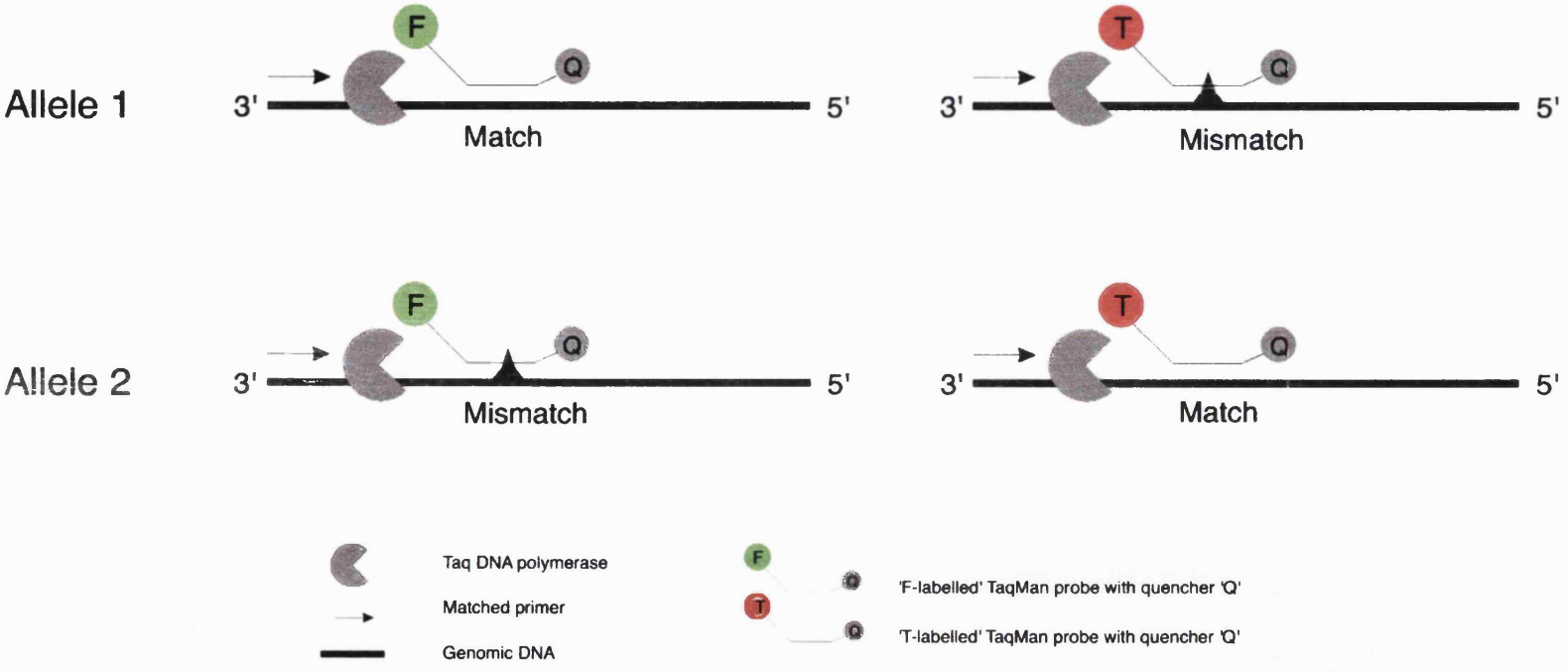


Fig 5.2: Allelic Discrimination using fluorogenic probes and the 5' nuclease assay. Increase in FAM signal equals homozygous for Allele 1; Increase in TET Signal equals homozygous for Allele 2; increase in FAM and TET signal equals homozygous for Alleles 1 and 2.

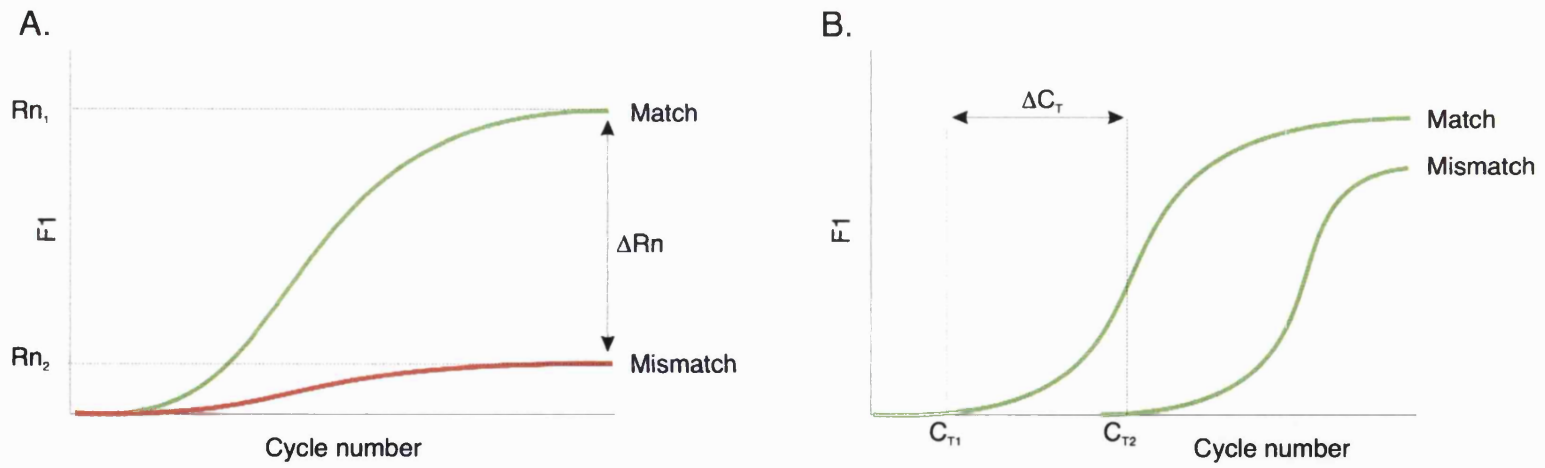


Figure 5.3 Typical read-outs. [A] *Conventional allelic discrimination:* Two allele-specific TaqMan® probes are incorporated, designed to anneal to the SNP site. TaqMan® probes are differentially labelled such that a specific signal can be assigned to a particular genotype. The signal giving the largest change in signal is considered the most efficient reaction, and is used to assign the genotype (Rn1-Rn2 = ΔRn). [B] *Combined TaqMan® and allele-specific PCR strategies*: A single probe is included in the reaction, annealing to a sequence within an allele-specific amplicon (not spanning the SNP site). Two separate reactions are performed containing different allele-specific primers and the same TaqMan® probe. A mismatch is shown to have a higher C_T than a matched allele-specific PCR reaction, so the differences in cycle threshold are used as a measure of PCR efficiency (C_T1- C_T 2= Δ C_T).

Consequently, several reports have been published that describe the combination of the TaqMan® assay with allele-specific amplification. These include TaqMAMA (Glaab and Skopek, 1999), TaqMan-ASA (Fujii *et al.,* 2000), allele-specific TaqMan PCR assay (Mizugaki *et al.,* 2000; Matsubara *et al.,* 1999). In each case, a single "generic" probe is included in a PCR directed by an allele-specific primer and a common primer. The probe is designed such that it is upstream of the allele-specific primer, annealing between the two amplification primers. Following two separate tests incorporating each different allele-specific primer with the same probe, the genotype is inferred by comparing C_T values (Figure 5.3B), as opposed to the ΔRn levels evaluated in the conventional system. The theory of this analysis is that a mismatched allele-specific primer generally gives a signal above threshold in later cycles than a matched test (i.e. it has a higher C_T value than a match), attributed to a decrease in priming ability and PCR efficiency (Glaab and Skopek, 1999). As long as the C_T value for a match is reproducibly lower than a mismatch, genotyping is possible in these systems.

Each 'combined' method described above has incorporated the discriminatory advantages of allele-specific PCR into the test, at a significant cost of returning the bi-allelic genotyping to a two-tube format. In a clinical assay where driving down genotyping costs is key, a two-tube system requiring double the amount of reagents than a single tube assay, and an internal control amplification, may not be favourable. This chapter describes an enhancement to the conventional allelic discrimination formats and allele-specific TaqMan® assays by adapting single tube bi-directional PCR (Figure 3.1, pg. 81) with allele-specific amplicon detection using TaqMan® probes.

5.1 Results

5.1.1 Assay design

A schematic of the bi-directional PCR/ TaqMan® probe assay is shown in Figure 5.4, using the A→T mutation in codon 6 of the β-globin gene as a SNP target (Section 1.1). Two primer sets were included in the same single-tube PCR test; one designed to amplify an 86 bp amplicon in the presence of the wild type allele (SC_WT_AS/ SC_WT_OP), and a second designed to amplify a 112 bp product in the presence of a mutant allele (SC_MUT_AS/ SC_MUT_OP). SC_WT_AS has an 'A' nucleotide at the 3' end matching the 'T' base on the coding strand of the normal b-globin gene. SC_MUT_AS has an 'A' nucleotide at the 3' end, matching the 'T' base on the coding strand side of the sickle cell mutation. Each primer was designed with a similar T_m to allow generic cycling conditions.

Two fluorescent-labelled TaqMan® probes (Appendix I), WT_FAM and MUT_JOE, were designed according to manufacturers' guidelines (Table 5.1) to lie as close as possible to the 3' end of each corresponding allele-specific primer surrounding the SNP site. The wild type probe (WT_FAM) was designed as complementary to the sequence upstream of the wild type allele-specific primer (SC_WT_AS). FAM was covalently linked to the 5' end of the WT_FAM probe, quenched by TAMRA attached via linker arm located at the 3' end of each probe. The mutant probe (MUT_JOE) was designed as complementary to the sequence upstream of the mutant allele-specific primer (SC_MUT_AS). JOE was covalently linked to the 5' end of the MUT_JOE probe, again quenched by TAMRA. A phosphate group was attached to the 3' end of both probes prevented extension by *Taq* polymerase during PCR (Holland *et al.,* 1991).

Sequences were checked using Primer Express Software (Applied Biosystems) and synthesised by Oswel Research Products (Southampton, UK). Figure 5.5 displays the sequence of interest with primer and probe alignments.

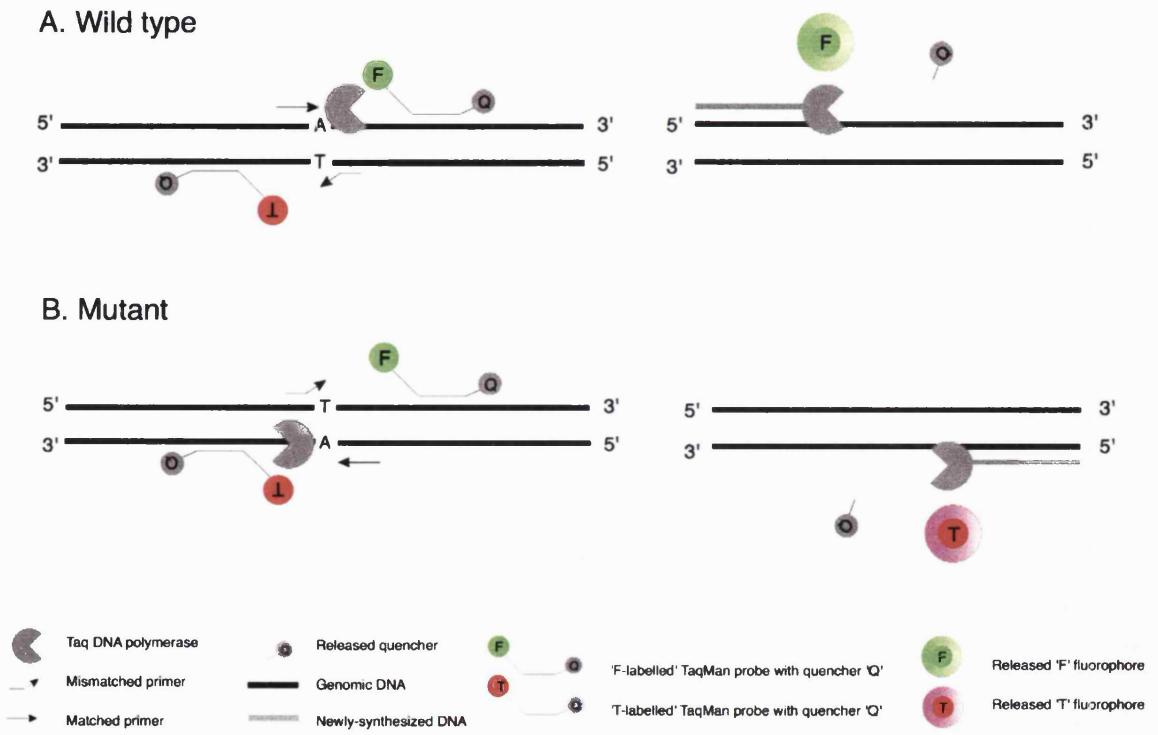


Figure 5.4. A schematic of the combined bi-directional TaqMan® system. The sequence specific probe is designed to anneal upstream of a matched allele-specific primer. Two probes and two primer sets are included in the reaction; both designed to lie in a bi-directional orientation around the SNP site. Therefore the probes will anneal to different strands of the DNA. A matched allele-specific primer will be extended, and consequently the corresponding probe will be cleaved to generate a detectable signal at the appropriate wavelength. Thus, the fluorophore giving the maximal signal is used to assign the genotype.

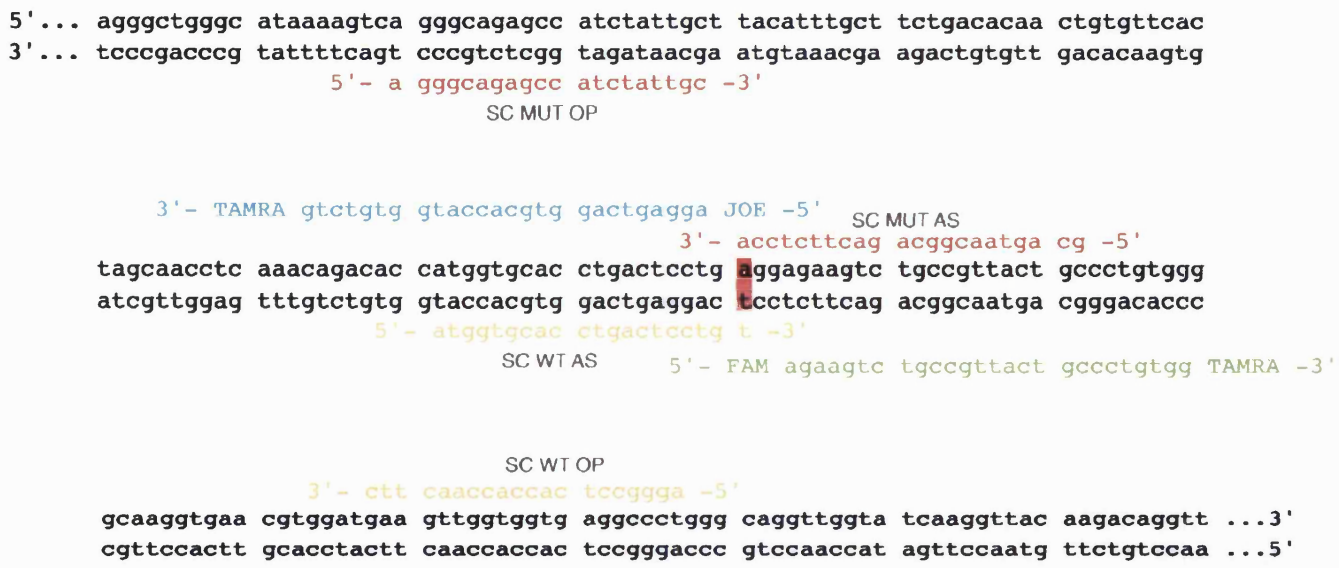


Figure 5.5: Sequence of interest with probe alignments. SC_WT_AS has an 'A' nucleotide at the 3' end matching the 'T' base on the coding strand of the normal b-globin gene. SC_MUT_AS has an 'A' nucleotide at the 3' end, matching the 'T' base on the coding strand side of the sickle cell mutation.

Probe and primer design recommendations for TaqMan® assays on the ABI 7700 Prism SDS
Amplicons should designed within 50-150 bp range
Primer T_m should be 58-60°C
G/C content of probe should be within the 20-80% range
Avoid runs of an identical nucleotide, especially guanine where runs of 4 or more G's should be avoided
Probe T_m should be 68-70°C
Probe should not have a G on the 5' end
For probe annealing, select the strand that gives the probe more C than G bases

Table 5.1: Recommendations for designing TaqMan® probes (Applied Biosystems Technical Bulletin, 2001).

5.1.2 Assay optimisation and PCR amplifications

The combined TaqMan® and bi-directional PCR assays were performed using an ABI Prism 7700 Sequence Detection System (Applied Biosystems) according to manufacturer's instructions. All reactions were prepared in a 96-well plate, using a TaqMan® Universal Master Mix (containing AmpliTaq® Gold DNA polymerase, AmpErase® UNG, dNTPs with dUTP, a passive reference dye, and optimised buffer components in a proprietary formulation; Applied Biosystems), oligonucleotides, 250 nM each labelled probe and 5-25 ng human genomic DNA.

Primer titration was used to define primer concentrations that generated allele-specific amplicons according to the genotype of the template DNA. Given that the Mastermix contains optimised concentrations of all components apart from primers and probe, only a simple primer titration was necessary to define discriminatory conditions. Two sets of nine reactions were performed in triplicate, incorporating either wild type or mutant control DNA template. Primers were aliquoted according to Table 5.2 wherein the range of concentrations was decided empirically. The distribution of samples in the 96-well format is shown in Figure 5.6.

Samples were taken through a 2 min hold at 50°C for optimal AmpErase® Uracil-N-glycosylase (UNG) activity, followed by a 10 min hold at 95°C for AmpliTaq Gold® DNA polymerase activation. This was followed by a cycling program of 95 °C for 15 s, 60 °C for 30s, for a total of 40 cycles. The fluorescence emission caused by cleavage of the probe upstream of an extended (matched) allele-specific primer was detected over the course of the run. For detection of the FAM signal, readings were captured within the emission wavelength range including 535 nm. For JOE, this range included 557 nm (Molecular Probes Handbook, 2001). The ABI Prism® 7700 SDS software automatically normalised the fluorescence signal against non-PCR related fluctuations by dividing the emission intensity of the reporter dyes by the emission of the passive reference dye (i.e. ROX, 6-carboxy-X-rhodamine) included in the reaction Mastermix. Samples were run on an agarose gel for clarification of fragment sizes according to Section 2.5.4.

Figure 5.7 shows a typical set of gel image data from the primer titration sets. Table 5.3 displays corresponding ΔRn data for wild type and Mutant DNA template. The ΔΔRn data represents the difference in the ΔRn for each probe in the presence of wild type or mutant template. In the presence of wild type DNA, a large signal corresponding to the FAM probe was detected above the instrument threshold. Concurrently, the signal from the JOE probe did not rise above the threshold level. In the presence of mutant DNA, a large signal corresponding to the JOE probe was detected above a threshold, and a signal was also detected above threshold from FAM at a significantly lower ΔRn than wild type tests. The levels of detectable fluorescence varied according to the primer concentrations included in the test.

Rxn	*WT AS, µM*	*WT OP, µM*	*MUT AS, µM*	*MUT OP, µM*
1	A	A	A	A
2	A	A	B	B
3	A	A	C	C
4	B	B	A	A
5	B	B	B	B
6	B	B	C	C
7	C	C	A	A
8	C	C	B	B
9	C	C	C	C

Table 5.2: Primer titration matrix. Primers were aliquoted according to this matrix, where A= 0.1 µM, B= 0.2 µM and C= 0.6 µM. Inter-set primer concentrations were chosen empirically to cover a wide range of low concentrations.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Rep 1 WT 1	Rep 1 WT 2	Rep 1 WT 3	Rep 1 WT 4	Rep 1 WT 5	Rep 1 WT 6	Rep 1 WT 7	Rep 1 WT 8	Rep 1 WT 9	0	0	0
B	Rep 2 WT 1	Rep 2 WT 2	Rep 2 WT 3	Rep 2 WT 4	Rep 2 WT 5	Rep 2 WT 6	Rep 2 WT 7	Rep 2 WT 8	Rep 2 WT 9	0	0	0
C	Rep 3 WT 1	Rep 3 WT 2	Rep 3 WT 3	Rep 3 WT 4	Rep 3 WT 5	Rep 3 WT 6	Rep 3 WT 7	Rep 3 WT 8	Rep 3 WT 9	0	0	0
D	Rep 1 MUT 1	Rep 1 MUT 2	Rep 1 MUT 3	Rep 1 MUT 4	Rep 1 MUT 5	Rep 1 MUT 6	Rep 1 MUT 7	Rep 1 MUT 8	Rep 1 MUT 9	0	0	0
E	Rep 2 MUT 1	Rep 2 MUT 2	Rep 2 MUT 3	Rep 2 MUT 4	Rep 2 MUT 5	Rep 2 MUT 6	Rep 2 MUT 7	Rep 2 MUT 8	Rep 2 MUT 9	0	0	0
F	Rep 3 MUT 1	Rep 3 MUT 2	Rep 3 MUT 3	Rep 3 MUT 4	Rep 3 MUT 5	Rep 3 MUT 6	Rep 3 MUT 7	Rep 3 MUT 8	Rep 3 MUT 9	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0
H	Ntc 1	Ntc 2	Ntc 3	Ntc 4	Ntc 5	Ntc 6	Ntc 7	Ntc 8	Ntc 9	0	0	0

Figure 5.6: 96- well format for optimisation reactions. A zero indicates that wells were left empty. 'Rep' indicates that this sample is a repetition in the triplicate. WT cells contain control wild type DNA. MUT cells contain control mutant DNA. Ntc indicates a no template control in which all reagents are present except for template.

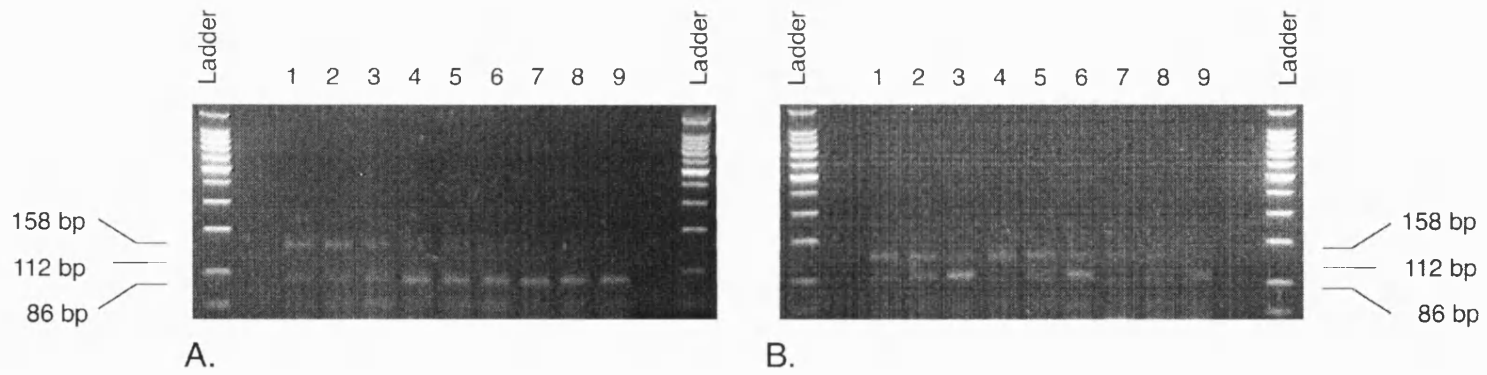


Figure 5.7: Gel image data from optimisation sets. Note lower efficiency in these amplifications including a probe. Levels of amplification were sufficient for fluorescence detection.

[A] Wild Type DNA

ΔRn	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
FAM	1.29±0.03	0.95±0.15	0.85±0.10	1.82±0.12	1.36±0.01	0.99±0.02	1.91±0.09	1.57±0.02	0.96±0.09
JOE	0	0	0	0	0	0	0	0	0
ΔΔRn FAM	0.87	0.61	0.61	1.43	1.05	0.79	1.67	1.36	0.82

[B] Mutant DNA

ΔRn	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
FAM	0.42±0.03	0.34±0.01	0.24±0.04	0.39±0.02	0.31±0.02	0.20±0.02	0.24±0.02	0.21±0.02	0.14±0.01
JOE	0.20±0.01	0.24±0.02	0.27±0.02	0.17±0.02	0.20±0.01	0.22±0.00	0.11±0.01	0.13±0.00	0.15±0.02
ΔΔRn JOE	0.20	0.24	0.27	0.17	0.20	0.22	0.11	0.13	0.15

Table 5.3. Tables display typical fluorescence readings (ΔRn) from FAM and JOE probes during assay optimisation.

5.1.3 SNP genotyping

The conditions of Reaction 6, Table 5.2 were used to genotype twenty-two randomised samples and a single control sample void of template, since both probes gave a large ΔRn value (Figure 5.8). Eight samples gave a positive FAM signal only (Figure 5.9A), six samples gave a positive JOE signal with a very low signal from FAM (Figure 5.9B), and three samples gave a measurable fluorescence increase from both reporter dyes (Figure 5.9C). Six samples, including the no template control, did not generate a measurable fluorescence signal.

The ΔRn value from the FAM signal (WT allele) was plotted vs. the ΔRn value from the JOE signal (MUT allele). Four distinct cluster regions were present depending upon genotype (Figure 5.10). Eight samples were clustered around the *y*-axis (FAM signal), six samples were clustered around the *x*-axis (JOE signal), and three samples were placed near to the *x*=*y*-axis resulting from signals from both dyes. Six samples formed a cluster around the origin where no measurable fluorescence was detected.

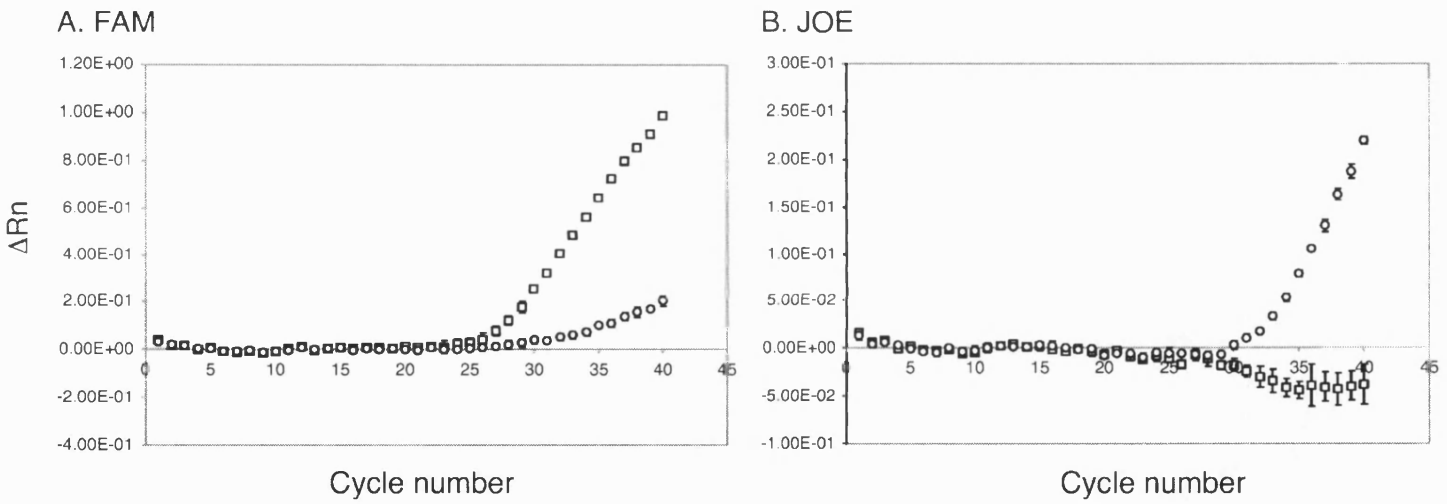


Figure 5.8: Typical data set showing [A] FAM and [B] JOE signals in the presence of wild type DNA ( ) or mutant DNA (o) under the conditions defined by Reaction 6, Table 5.2. Reactions were run in triplicate and error bars are shown. The negative JOE signal observed in [B] is attributable to the baseline correction used in data analysis (see Figure 1.18B, pg. 59).

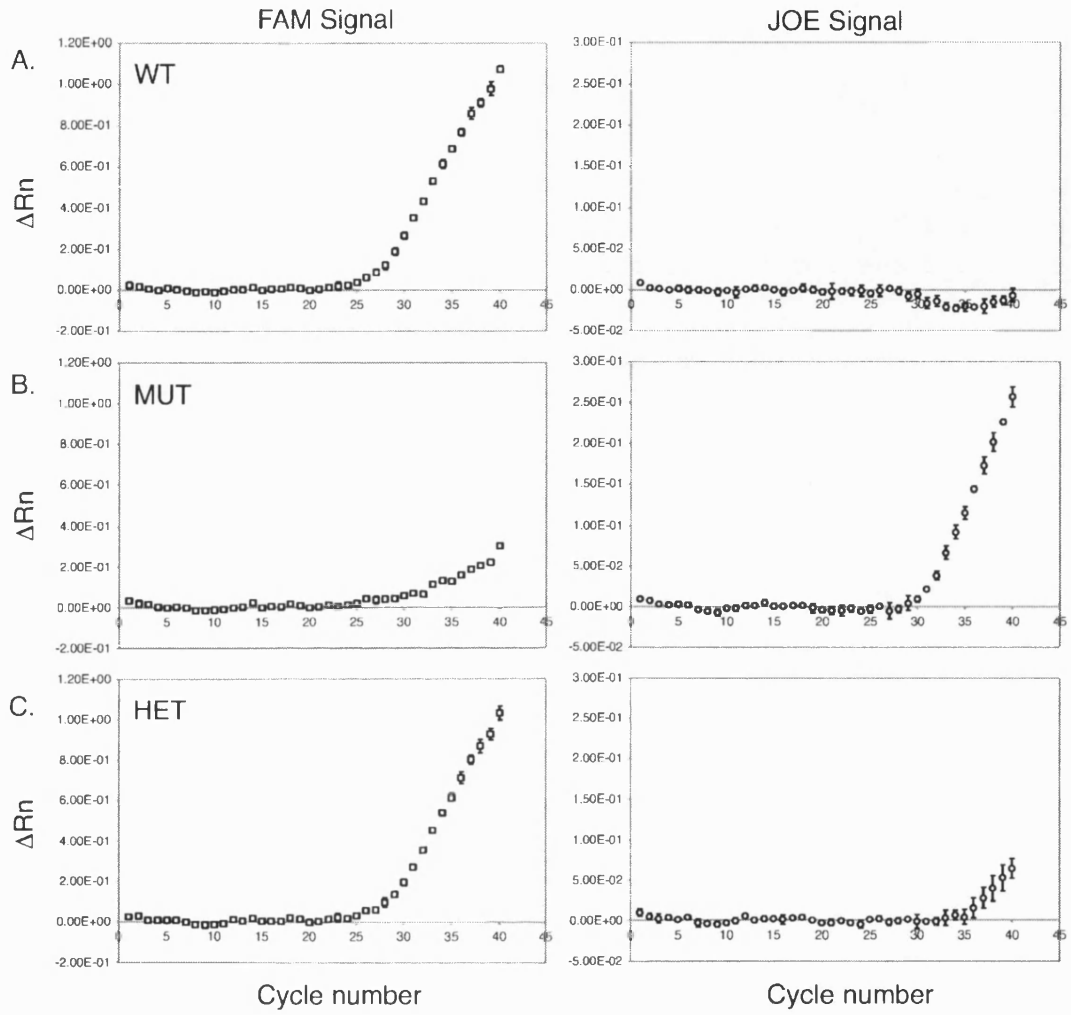


Figure 5.9: Typical fluorescence signals from FAM (left hand side) and JOE (right hand side) probes in the presence of [A] wild type, [B] mutant and [C] heterozygous DNA under the conditions of Reaction 6, Table 5.2.

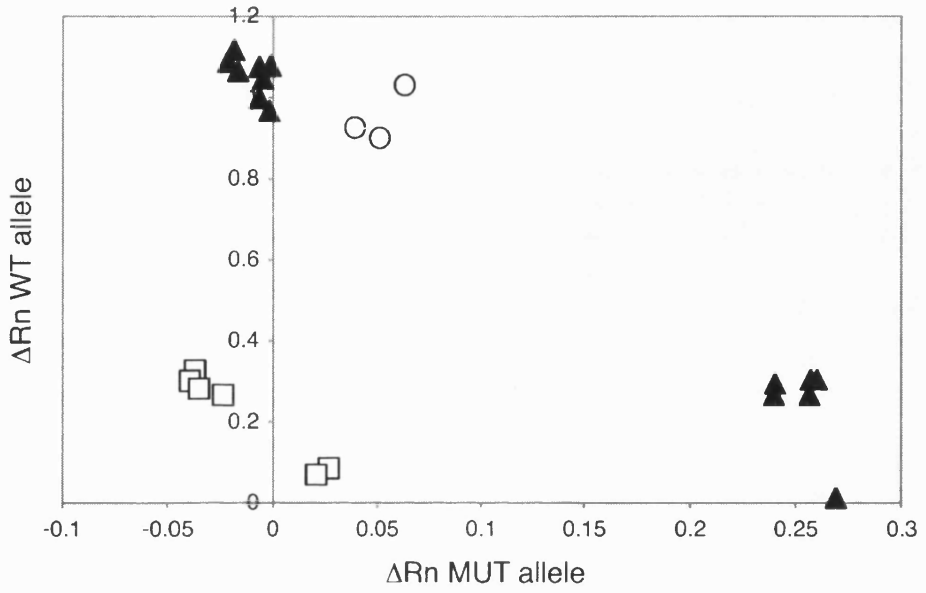


Figure 5.10: SNP calling. Points lying nearest to the x-axis show homozygous mutant inheritance of the mutation. Points lying near the y-axis show homozygous wild type DNA, and the three points (o) lying near to the $x = y$ axis show heterozygous inheritance. Points located round the origin were control samples void of template, or mistypes.

5.2 Discussion

Combining the single tube bi-directional PCR format with fluorogenic probes for the specific detection of allele-specific products provides an interesting alternative to detecting single nucleotide changes in a DNA sequence. The test compares the amplification efficiency of two separate allele-specific amplicons, by incorporating two-sequence specific probes into a single test. An internal control is inherent since any successful PCR will generate an allele-specific amplicon. A simple 'yes-no' result is provided, which is ideal in a clinical setting where ambiguity should be eliminated to prevent misdiagnosis.

A primer titration was performed to reveal optimal conditions that gave allele-specific amplification without artefacts (Figure 5.7). The presence of the well-characterised sickle cell anaemia mutation was detected in a small trial of twenty-two human DNA samples, according to the magnitude of change in normalised fluorescence per reaction for each probe. Samples homozygous for the wild type allele gave a large ΔRn with the FAM detection probe, and a corresponding negligible JOE signal (Figure 5.9A). Samples homozygous for the mutant allele gave a large ΔRn with the JOE probe, and a corresponding negligible FAM signal (Figure 5.9B). Heterozygotes were identified by a signal from both reporter dyes, slightly lower than for a homozygote due to molar ratios of each allele (Figure 5.9C). Where a signal was detected above baseline, concordance with RFLP data was 100% (Figure 4.19, pg. 153). Five samples failed to give a signal above baseline, likely due to insufficient starting DNA concentration.

The combined assay has increased flexibility in probe design compared to the conventional method because the target region for probe alignments is not restricted to the area containing the SNP site. Rather, any part of the sequence upstream of either allele specific primer is accessible to probe design. Primer design is also unrestricted since wild type and mutant primer sets can be designed without considering T_m constraints, as in fluorescent bi-directional PCR (Chapter four). Smaller amplicons (between 50-150 bp) are recommended, as with most

protocols used today. It would be beneficial to design amplicons of a similar size to minimise differences in amplification efficiency. Here, amplicons of 86 bp and 112 bp were shown to be suitable.

The most commonly used commercially available dyes for 'two-dye' set-ups in TaqMan® assays were 6-FAM and VIC (or JOE; based on an unpublished survey). This is due to the minimal spectral overlap between the dyes, so that separate signals can be attained simultaneously without significant cross talk, which would increase the error in the read outs from each probe. In these experiments, the magnitude of JOE signal was consistently lower than the FAM signal. Although the ABI Prism® 7700 SDS is able to excite a wide range of fluorophores using laser excitation (488 nm), FAM (excited at 495 nm) has a greater spectral overlap with the shortest excitation wavelength than JOE (excited at 525 nm). Consequently, the highest signal of excitation is detected from FAM, even if identical amounts of allele-specific amplicon are produced. The availability of different fluorophores combined with the ability of the instrument to monitor many dye layers makes possible the identification of at least two multiple targets in homogenous assays, utilised in the assay described here.

Other features of PCR instrumentation play a key role in the success and application of PCR-based assays. In particular, the throughput and speed of thermal cycling. Currently, real-time PCR instruments that offer integrated product amplification and detection, a wide excitation source for flexible multiplexing, rapid cycling, and greater than 32 samples per run are unavailable (Table 4.1, pg. 118). The combined assay described here would benefit from rapid cycling due to improved specificity using shorter extension times (Wittwer *et al.,* 1993) and increased throughput. However, the ABI Prism 7700 SDS is based on slow Peltier elements for temperature cycling and is limited to maximal temperature transitions of around $3°C.s^{-1}$. A truncated thermal cycling program was used here in order to lower the assay time reducing the overall time to two hours.

The LightCycler™ instrument uses rapid air exchange as the medium for heat transfer and is capable of significantly faster ramp rates ($20°C.s^{-1}$) such that the

same test could be completed within 30 minutes. The rate-limiting step would then become the minimum annealing time for effective 5' nuclease activity by *Taq*, typically in the order of 10 seconds (Holland *et al.,* 1991). However, the LightCycler™ instrument is limited to 32 tests per run and does not have the same flexibility of dye choice as the ABI 7700 Prism SDS, due to the fixed wavelengths of excitation and emission wavelength detection. The blue LED of the LightCycler™ is able to excite fluorophores at a single wavelength (470 nm). This can efficiently excite the FAM fluorophore due to sufficient spectral overlap (λ_{ex}495/λ_{em}535 nm), but is unable to excite JOE-labelled probes (λ_{ex}525/λ_{em}557 nm).

Specialised FRET probes such as wavelength shifting molecular beacons (Tyagi *et al.,* 2000), or FRET scorpion primers (Thelwell *et al.,* 2001) could be used to circumvent this restriction. These nucleic acid hybridisation probes fluoresce in a variety of different colours (depending upon the nature of the emitter fluorophore), yet are excited by a common monochromatic light source (Figure 5.11). The harvester fluorophore is chosen to absorb strongly in the wavelength range of the light source (i.e. FAM could be excited by the LightCycler™ light source) which, when excited will excite the emitter fluorophore, generating a characteristic detectable signal. The only drawback to these probes is the increased numbers of fluorophores required and complex design considerations, which ultimately increases cost.

The assay would also be compatible with other probe systems suited to rapid cycling. These include Resonsense® (Lee *et al.,* 2001), and HyBeacon® probes (French *et al.,* 2001) for example. Two probes labelled with different dyes, with complementarities to regions within the non-primer sequence of either allele-specific amplicon could be included in a single PCR test following the protocol described here. Of particular interest, Resonsense® probes use a DNA intercalater as the donor fluorescence. This would enable the excitation of dyes compatible with the LightCycler™ instrument by FRET mechanisms, increasing assay speed. Evaluating these other probe formats would be the next stage in development of this assay.

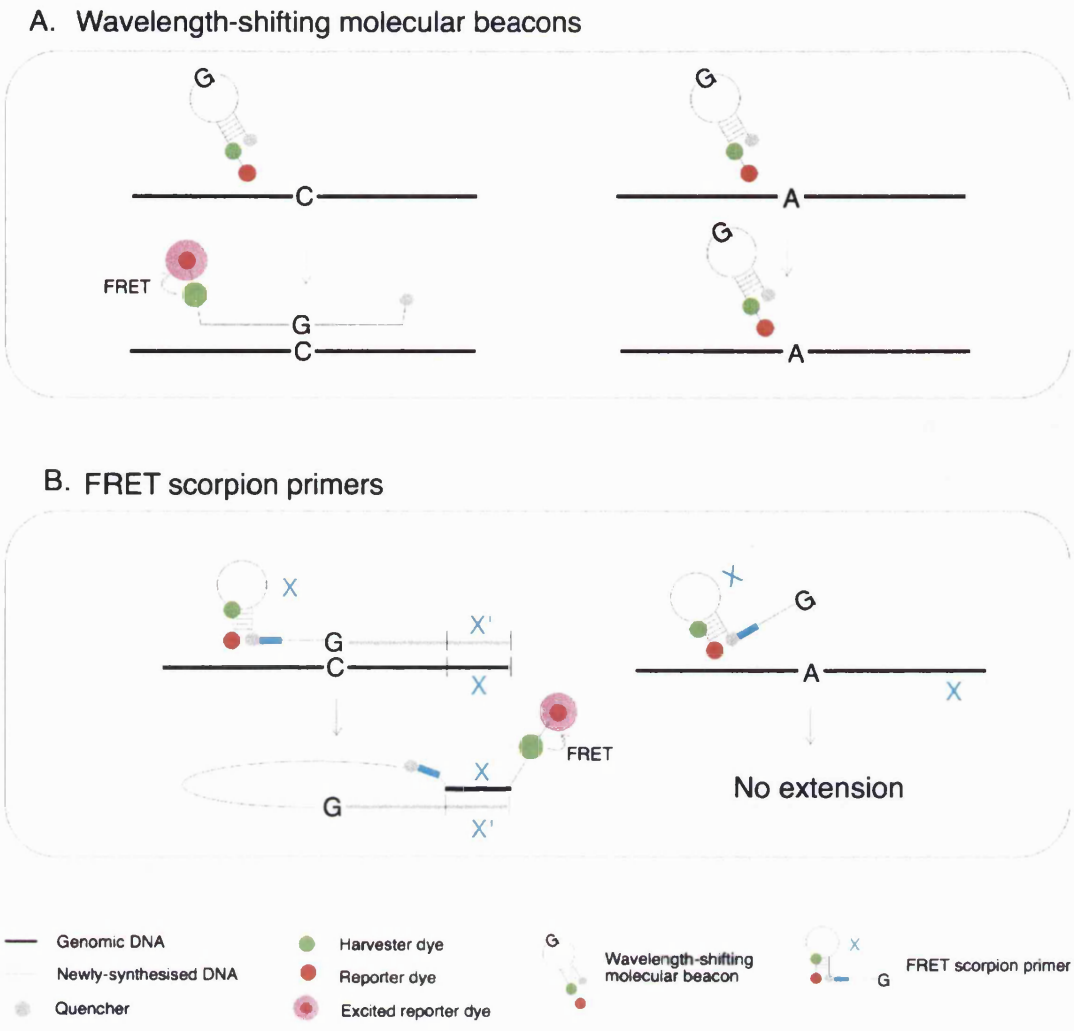


Figure 5.11: [A] *Wavelength-shifting molecular beacons*: Use the same principles as conventional molecular beacons except the beacons is labelled with two fluorophores (on the 5' arm) and one quencher (at the end of the 3' arm). A harvester fluorophore is at an internal location in the 5' arm of the beacon such that it is opposite to the quencher moiety. A reporter fluorophore lies at the distal end of the 5' arm. The harvester fluorophore is chosen such that it can be excited by a monochromatic light source, and its emission wavelength should have spectral overlap with the excitation wavelength of the reporter fluorophore. In the absence of a target sequence, probes are dark because the fluorescence from the harvester fluorophore is quenched. In the target-bound conformation, the energy absorbed by the harvester fluorophore is transferred by FRET to the reporter fluorophore, which then releases energy as light in its characteristic colour (Tyagi *et al.*, 2000). [B] *FRET scorpion primers*: In the closed form the quencher is able to quench fluorescence from the harvester fluorophore and reporter fluorophore. After extension of the scorpion primer the probe binds to the target separating the quencher from the two fluorophores. The harvester fluorophore absorbs energy from the light source and transfers the energy by FRET to the reporter fluorophore. Excitation of the reporter dye occurs and an increase in its characteristic is observed (Thelwell *et al.*, 2000).

It is noted that since completing this work (November 2001), researchers have addressed the issue of inefficient destabilisation in the presence of a mismatch. A chemical modification of the fluorogenic probe with a minor groove-binding agent is now used to increase duplex stability. This new generation of probe incorporates minor groove binder (MGB) groups, to form extremely stable duplexes with single-stranded DNA targets, allowing shorter probes to be used for hybridisation-based assays. The MGB probes were more specific for single base mismatches than standard DNA probes and produced better fluorescent signals. A larger destabilisation was seen with shorter probes when compared to longer conventional probes. In certain cases, the ΔT_m of a matched and mismatched probe was just 0.5°C, compared with a minimum ΔT_m of 10°C with the MGB probe (Epoch, 2001). The MGB probe is now commercially available, highlighting the current acceptability and continuing evolution of these probes for SNP detection.

To summarise, the test returns combined AS-PCR/TaqMan® probe methods to a single tube format. The current limitation of this assay is the high synthesis cost of probes modified with fluorescent and quenching moieties, compared to dsDNA intercalaters. These costs are particularly acute in high throughput applications. Nevertheless, this has not limited the number of assays that utilise sequence-specific probes (as shown in Chapter One) and arguably costs may be driven down through competition. The continued development of innovative software and algorithms may help to improve automation in probe design and selection based on raw sequence information and may reduce the upfront costs of assay design.

Chapter Six

Kinetic Characterisation of PCR: Application to primer mismatches and quantitative PCR

6.0 Introduction

Until recently the polymerase chain reaction (Mullis *et al.,* 1986; Saiki *et al.,* 1985) has predominantly been used as a binary tool to indicate the presence or absence of a particular target sequence. Instruments capable of quantitative detection of amplification in real time (qPCR; Table 4.1, pg. 118) have enabled sensitive analysis of gene expression and accurate determination of copy number (Bustin, 2000; Wittwer *et al.,* 1997), enhancing the utility of PCR as a diagnostic.

Despite the routine use of PCR in molecular biology, enzymatic and kinetic studies have been limited. Early characterisation studies detailed the efficiency with which *Taq* DNA polymerase differentiates between matched and mismatched 3' nucleotides (Kuchta *et al.,* 1987; Petruska *et al.,* 1988; Mendelman *et al.,* 1990; Wong *et al.,* 1991; Christopherson *et al.,* 1997). These considered the action of *Taq* DNA polymerase in isolation of PCR dynamics, assessing the relative extension efficiencies in the presence of a single dNTP species. This significantly underestimates the kinetic effects known to influence reaction fidelity and efficiency, and limits application in the design of genotyping tests based on allele-specific PCR (AS-PCR).

Subsequent studies, summarised earlier (Table 3.2; pg. 87), considered the effects of primer 3' nucleotide mismatches on primer extension by *Taq* DNA polymerase; primarily using end point amplicon yields as a measure of reaction efficiency during AS-PCR. The results exhibited large variability since the analysis used different reaction conditions and components, known to influence reaction specificity (Table 3.3; pg. 88). Significantly, previous studies used slow Peltier-based thermal cyclers that have slow temperature transitions rates. The rapid cycling and short extension

times of contemporary qPCR instruments has been shown to strongly disfavour mismatch extension (Wittwer *et al.,* 1993), suggesting that abrupt transitions in temperature have effects that have not been previously considered.

Michaelis-Menten kinetics have been used to describe the variation of the rates of many enzyme-catalysed reactions as the substrate or effector concentration is varied (Wharton and Eisenthal, 1981). There is experimental evidence from previous studies to show that DNA polymerase reactions obey this typical behaviour (Schnell and Mendoza, 1997; Goodman *et al.,* 1993). Therefore, the Michaelis-Menten system can be utilised to underlie the principles of a theoretical description of the polymerase chain reaction and obtain quantitative predictions of enzyme characteristics including catalytic activity, effects of specific inhibitors, and an enzyme's specificity toward a particular substrate (Cornish-Bowden, 1979).

In a steady state (referring to the situation in which the value of a particular quantity is constant), and for enzyme concentrations negligible to those of the substrate, the formation and breakdown of an enzyme-substrate complex is described by the relation,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (6.1)$$

where $[E]$, $[S]$, $[ES]$ and $[P]$ represent the concentrations of free enzyme, free substrate, enzyme-substrate complex and product, respectively.

The rate of product formation or initial velocity is then given by the Michaelis-Menten equation:

$$v = \frac{V_{\max}[S]}{[S] + K_M} \qquad (6.2)$$

The Michaelis-Menten rate constant, K_M, is given by:

$$K_M = \frac{k_2 + k_{-1}}{k_1} \tag{6.3}$$

and the maximum velocity by

$$V_{\max} = k_2[E_t] \tag{6.4}$$

where $[E_T]$ represents the total enzyme concentration (Michaelis and Menten, 1913). In this simple description of $V_{\max}$, k_2 is identical to the *catalytic constant*, k_{cat}, a fundamental quantity defining the number of substrate conversions the enzyme can catalyse in unit time (Cornish-Bowden, 1979).

The Michaelis-Menten equation (equation 6.2), predicts simple hyperbolic behaviour or saturation kinetics (Figure 6.1 A). At sufficiently low concentrations of substrate, $[S]$, v increases linearly with $[S]$. But as $[S]$ is increased, this relationship begins to break down and v increases less rapidly than $[S]$ until, at sufficiently high or saturating $[S]$, v tends toward the limiting value $V_{\max}$ (Equation 6.4).

For a number of different reasons, enzymes often show departure from the behaviour predicted by the Michaelis-Menten equation. One cause is high substrate inhibition, in which the formation of abortive complexes caused by an excess of substrate inhibits the enzyme. High substrate inhibition leads to a kinetic plot such as that shown in Figure 6.1 B. The curve is described by the equation,

$$v = \frac{V_{\max}}{\left[1 + \left(K_M / S\right) + \left(S / K_i\right)\right]} \tag{6.5}$$

Figure 6.1: High substrate inhibition. [A] A typical Michaelis-Menten plot showing simple hyperbolic behaviour described by equation 6.2; [B] A typical Michaelis-Menten plot showing high substrate inhibition described by equation 6.5. Plot shows a deviation at higher substrate concentrations from the typical behaviour shown in [A]. Based on a figure by Eisenthal and Danson, 2002.

where V_{max} is the maximum velocity, K_M the Michaelis-Menten constant, and K_i the inhibition constant. In this formulation an additional term, K_i, is introduced into the denominator of the simple Michaelis-Menten equation; this term becomes predominant as the substrate concentration increases, resulting in inhibition of the observed rate (Wharton and Eisenthal, 1981). For example, at low substrate concentration [S], (S/K_i) is insignificant compared to $1+(K_M/S)$, and equation 6.5 becomes identical to the classic Michaelis-Menten equation. It follows that as S increases, the value of (S/K_i) increases, describing the portion of the curve where the rate of reaction drops rapidly (Wharton and Eisenthal, 1981).

High substrate inhibition is regularly cited as a main cause of the plateau phase in PCR. Therefore it is possible that the kinetics of PCR may exhibit the characteristics described by equation 6.5. During this chapter, the development of a kinetic model for PCR based on the equation for high substrate inhibition in Michaelis-Menten systems is described. The kinetic model is applied to rapid cycle AS-PCR (Newton *et al.,* 1989), a method used to discriminate between alleles of a gene based on single base pair differences (see Chapter three). The model provided a method by which Michaelis-Menten parameters could be estimated, and the likelihood of misextension for a particular polymorphism could be inferred by a quantitative parameter.

The kinetic model was also used in the development and validation of a novel method for the quantification of nucleic acid using real-time quantitative PCR. Previously, theoretical descriptions of PCR have been put forward that apply different mathematical approaches to simulate various physical parameters of the system, mainly for quantitative-competitive PCR application. These include statistical estimations of amplification rate (Peccoud and Jacob, 1998), probability of DNA replication (Stolovitzky and Cecchi, 1996), probability of DNA binding rates (Velikanov and Kapral, 1999), and derivation of expressions for amplification efficiency (Schnell and Mendoza, 1997A; Schnell and Mendoza, 1997B; Raeymaekers, 1993; Weiss and Haeseler, 1997). However, each of these models has yet to be evaluated using real PCR data.

6.1 Results

6.1.1 *Allele-specific PCR Amplifications*

PM_OP_98 and PM_AS_X (where X is G, C, A or T) were used to direct the amplification of a 98 bp allele specific amplicon from a pGEM-T® Easy Vector-based plasmid DNA template containing a G, C, A or T variant nucleotide (prepared according to Chapter Two). All twelve possible SNP combinations (G→C, G→A, G→T, C→G, C→A, C→T, A→G, A→C, A→T, T→G, T→C, T→A) were tested in triplicate with a single control sample void of template.

PCR cycling and melt analysis protocols were performed on the LightCycler™ PCR instrument according to Section 2.0.2 except the annealing temperature was set to 65 °C for high stringency. Critically, each PCR cycle was monitored by continual fluorescence acquisition throughout. Reaction specificity was confirmed by executing a melt curve analysis immediately after amplification according to Section 2.0.3. In each case, a single PCR product was formed with a T_m of approximately 85.0 °C. No amplification was observed in samples void of template. Continuous monitoring data were exported into Microsoft Excel 2000 (Macintosh Version 9.1, Microsoft Corporation, USA) for further analysis.

6.1.2 *Collection of 'per cycle' rate data*

The reaction rate v was determined from the maximal increase in fluorescence signal $F1$ with respect to time, t, at 72°C at each cycle. The rates were estimated from the linear portion of the progress curve, as shown in Figure 6.2 C, according to the linear equation,

$$F1 = vt + c, \qquad (6.6)$$

where c is the $F1$ reading at time zero. The slope v of fluorescence signal increase was assumed proportional to the rate of nucleotide incorporation by *Taq* DNA polymerase since this will be concomitant with the intercalation of SYBR Green I molecules into the newly-synthesised double-stranded DNA.

Data points in Figures 6.3 A-D show the cycle dependence of initial rate, v, using template with base variant 'T'. Figures 6.4 A-D, 6.5 A-D and 6.6 A-D display similar rate data using templates with base variants 'G', 'C' and 'A', respectively. During early and late cycles fluorescence changes were negligible. This is in contrast to cycles where exponential and linear amplification of target would be expected and where significant increases in fluorescence were observed from the real time data. The lag phase during early cycles was attributable to fluorescence signals that were below the detection limits of the optics in the LightCycler™. The length of the lag phase varied according to base complementation, and was generally less for a matched template compared to a mismatch. This lag phase was followed by a rapid increase in reaction velocity, which reached a maximum, then decreased as cycle number was increased further.

A kinetic model was developed based on the observed profile of these plots, described in detail during Section 6.2.1. The final kinetic model is shown below,

$$v = \frac{V_{\max}}{\left[1+\left(\frac{K_M}{(S+C)}\right)+\left(\frac{(S+C)}{K_i}\right)\right]} \qquad (6.7)$$

where,

$$S = \frac{1}{(1+a*\exp^{(-f*n)})} \qquad (6.8)$$

where v is the reaction rate, $V_{\max}$, K_M, and K_i are Michaelis-Menten parameters, n is cycle number, C is a compensation factor, and a and f are parameters that define the shape and position of a sigmoid curve (Seggern, 1993). Equation 6.7 was used in non-linear regression analysis (NL-REG Version 5.4; Phillip H. Sherrad), to fit real data from the AS-PCR tests (solid line in Figures 6.3-6.6 A-D inclusive) to estimate informative kinetic parameters for each base configuration (Table 6.1).

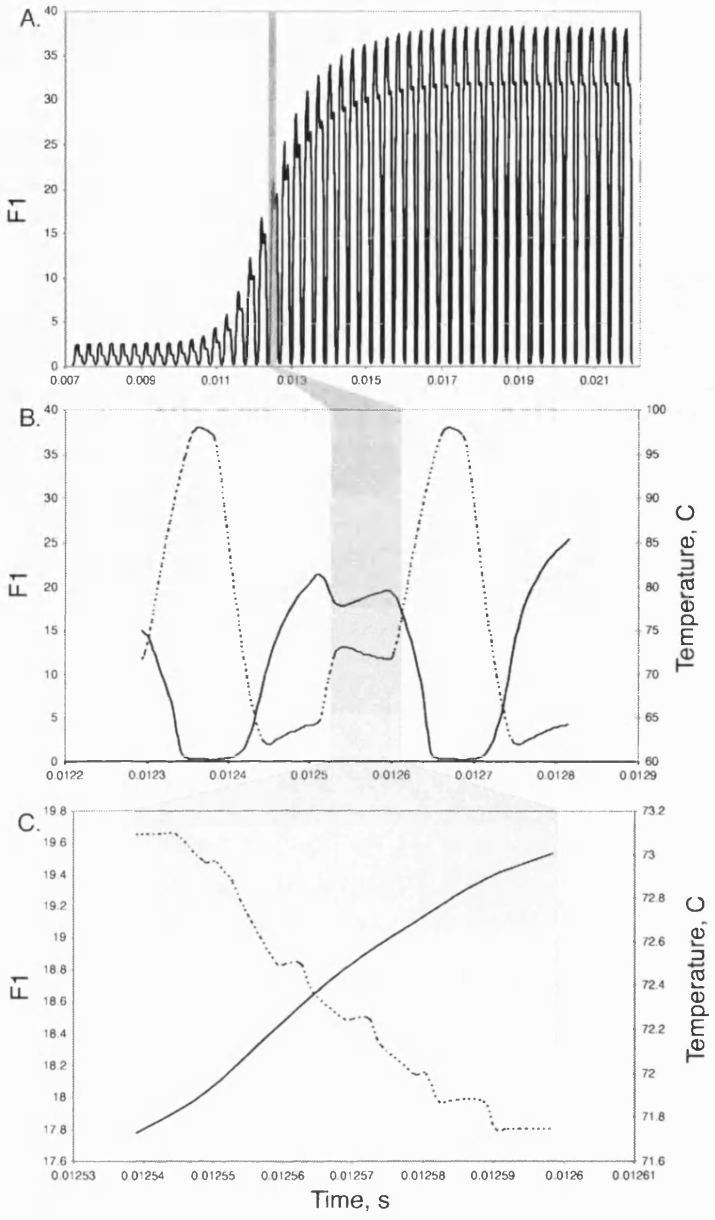


Figure 6.2: Grey area highlights section of interest from a matched allele-specific PCR test from template containing variant nucleotide T. [A] Continuous monitoring data over 50 PCR cycles. [B] Temperature dependence of product accumulation over a single cycle. [C] Product accumulation during the extension phase (~72°C) only.

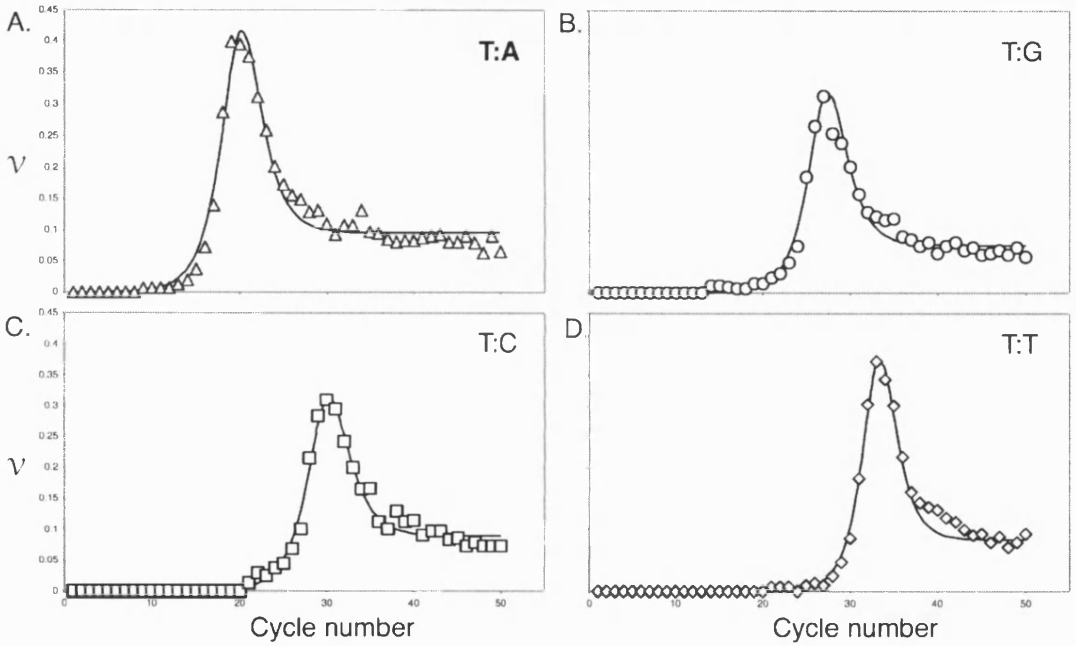


Figure 6.3: Typical graphs of rate v vs. cycle number for base combinations (template: primer) [A] T:A, [B] T:G, [C] T:C, and [D] T:T. The solid line represents the curve predicted by the model (Equation 6.7), and points represent data collected from real PCR tests using Equation 6.6. Correlation coefficient r^2 is >0.99 in each case.

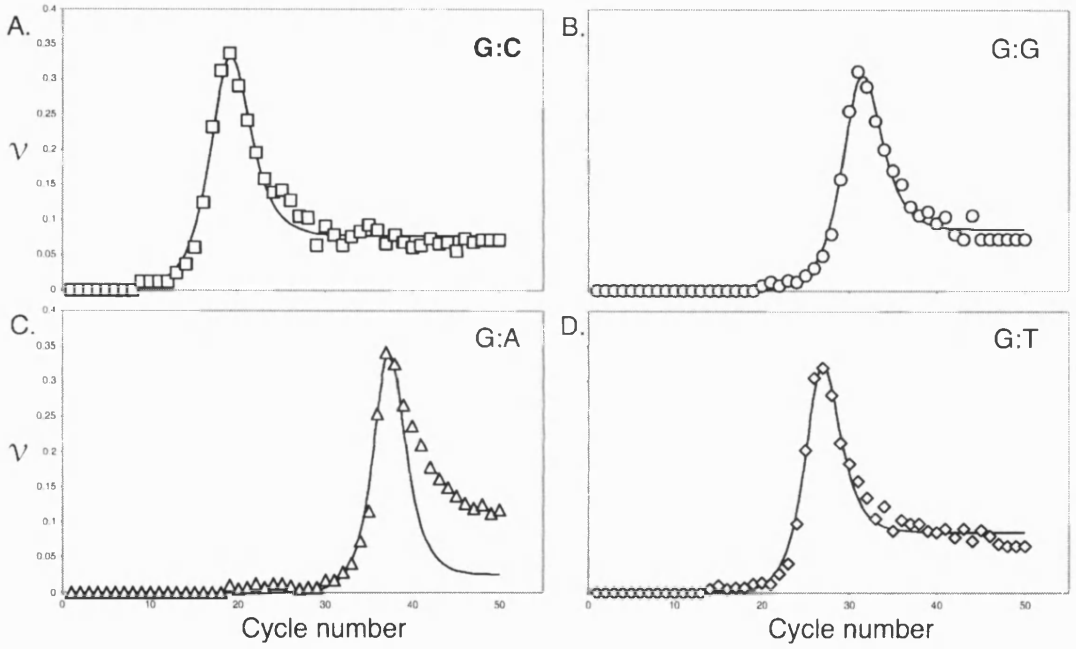


Figure 6.4: Typical graphs of rate v vs. cycle number for base combinations (template: primer) [A] G:C, [B] G:G, [C] G:A, and [D] G:T. The solid line represents the curve predicted by the model (Equation 6.7), and points are data collected from real PCR tests using Equation 6.6. Correlation coefficient r^2 is >0.88 in each case.

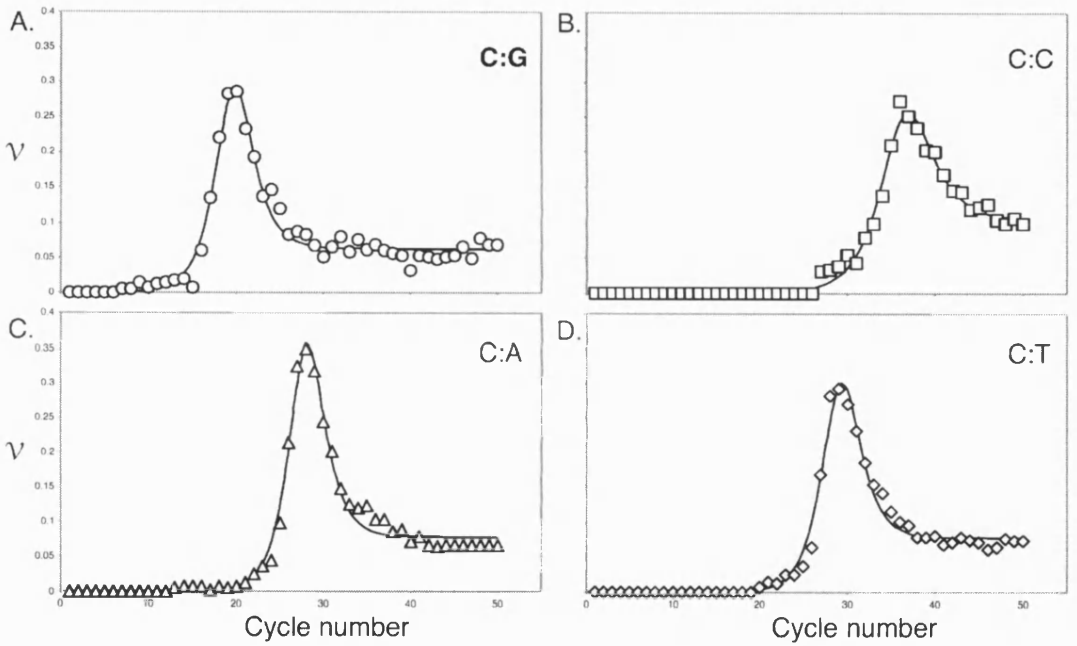


Figure 6.5: Typical graphs of rate v vs. cycle number for base combinations (template: primer) [A] C:G, [B] C:C, [C] C:A, and [D] C:T. The solid line represents the curve predicted by the model (Equation 6.7), and points are data collected from real PCR tests using Equation 6.6. Correlation coefficient r^2 is >0.99 in each case.

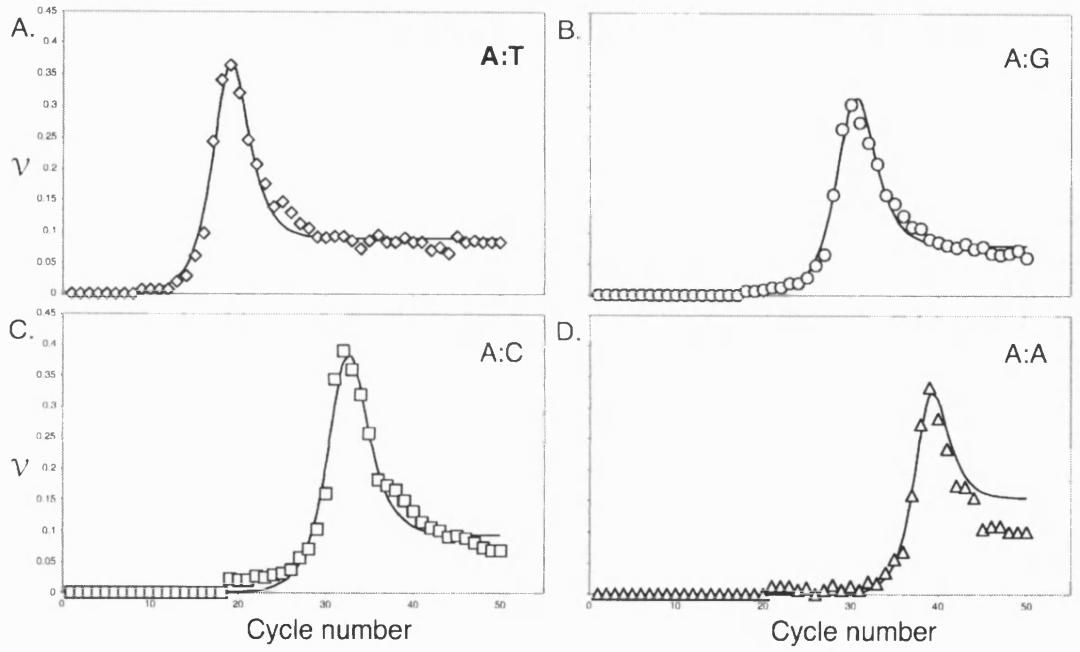


Figure 6.6: Typical graphs of rate v vs. cycle number for base combinations (template: primer) [A] A:T, [B] A:G, [C] A:C, and [D] A:A. The solid line represents the curve predicted by the model (Equation 6.7), and points are data collected from real PCR tests using Equation 6.6. Correlation coefficient r^2 is >0.98 in each case.

Template Base	*Primer 3' terminus*	V_{max}	K_M	K_i	V_{max}/K_M	[1]Ratio of V_{max}/K_M
	C	**530**	**1300**	**0.002**	**0.408**	**1.000**
G	G	448	43100	0.083	0.010	0.025
	A	88.1	479000	29.2	0.000	0.004
	T	302	23600	0.111	0.013	0.031
	C	290	14700	0.046	0.020	0.106
C	**G**	**335**	**1800**	**0.006**	**0.186**	**1.000**
	A	378	26100	0.093	0.015	0.078
	T	271	25400	0.128	0.011	0.058
	C	906	49900	0.037	0.018	0.089
A	G	402	32500	0.082	0.012	0.061
	A	45	1720000	369	0.000	0.000
	T	**377**	**1840**	**0.007**	**0.204**	**1.000**
	C	475	17700	0.031	0.027	0.120
T	G	669	16700	0.015	0.040	0.180
	A	**245**	**1090**	**0.013**	**0.224**	**1.000**
	T	160	116000	2.63	0.001	0.006

Table 6.1: Predicted kinetic parameters (shown to 3 s.f.) for each base configuration. Matched combinations are highlighted in bold type. Constants are displayed in arbitrary units.

[1] Ratio of specificity constants compared to a matched combination.

The ability of the model to predict substrate concentration at each cycle was tested by using the values estimated for the sigmoid parameters, a and f, to solve equation 6.8 in Microsoft Excel 2000 (Macintosh Version 9.1). The resulting sigmoid curve was plotted alongside the real time amplification data acquired at the end of each extension period at 72°C (Figures 6.7-6.10 A-D inclusive). The differences in actual and predicted fluorescence readings ($F1$) values are shown in Figure 6.11. Larger deviations were observed in mismatch-primed PCR tests.

6.1.3 Quantitative PCR amplifications

Four separate quantitation tests (Sets A – D) were performed using either pGEM-T® Easy Vector plasmid-based DNA (see Chapter Two) or human genomic DNA template to validate the kinetic model for qPCR application. Set A included amplification of a 223 bp fragment by primers ICF and ICR from plasmid DNA template. Set B included amplification of a 158 bp amplicon by primers SC_WT_OP and SC_MUT_OP from human genomic DNA, specifically the sequence surrounding the most common HbS sickle cell mutation. Set C included amplification of a 98 bp amplicon by primers WT_AS_C and WT_OP_98 from plasmid DNA (with variant nucleotide C). Set D included amplification of a 103 bp fragment by primers C282Y_WT_OP and C282Y_MUT_OP from human genomic DNA, specifically the sequence surrounding the C282Y mutation site conferring hereditary haemochromatosis.

PCR cycling and melt analysis protocols were performed on the LightCycler™ PCR instrument according to Section 2.0.2. Deviations from this protocol are summarised in Table 6.2. For each test, a 10x dilution series was run alongside a control test void of template, and three 'unknown' samples (Table 6.3). The 'unknowns' were selected to lie within the limits tested by external standards.

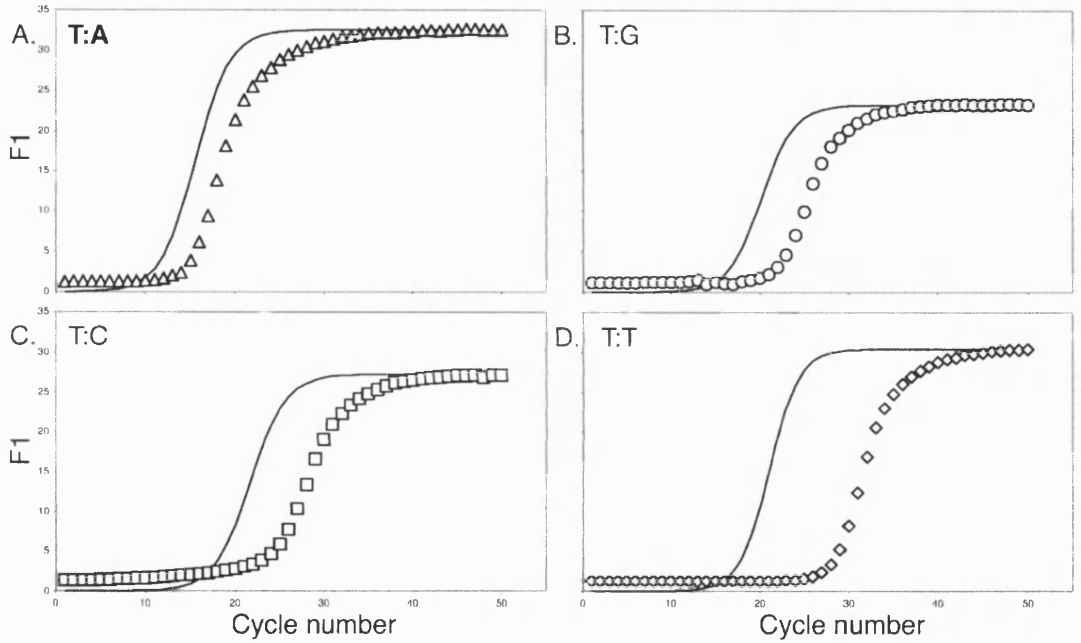


Figure 6.7: Correlation of real time substrate accumulation (data points) with that predicted by the sigmoid function (solid line). Graphs display base combinations (template: primer) [A] T:A, [B] T:G, [C] T:C, and [D] T:T.

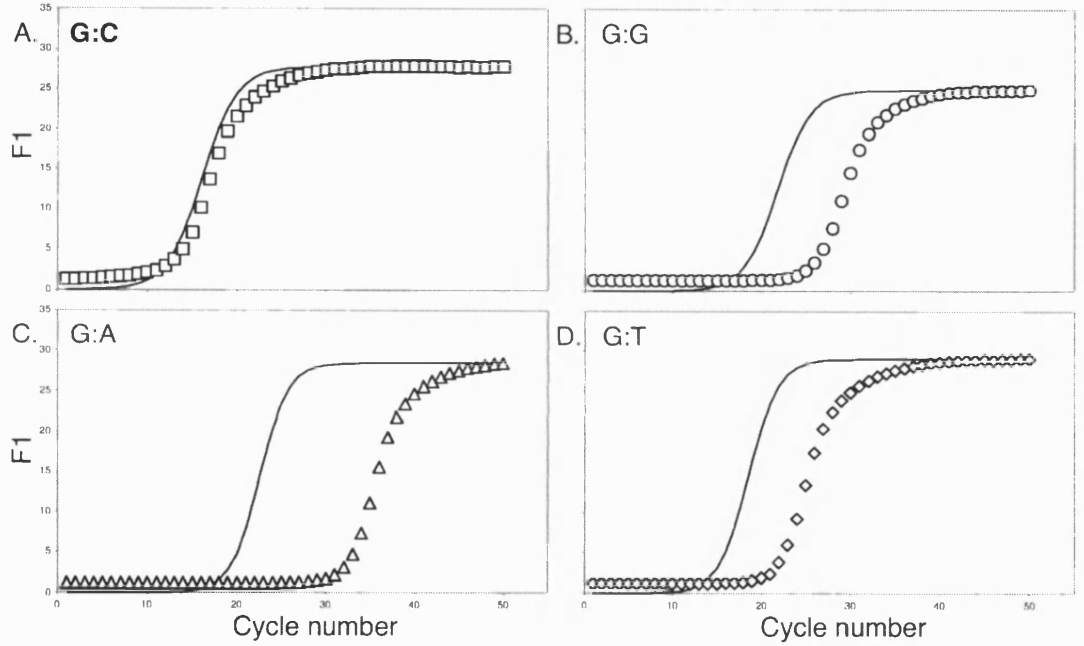


Figure 6.8: Correlation of real time substrate accumulation (data points) with that predicted by the sigmoid function (solid line). Graphs display base combinations (template: primer) [A] G:C, [B] G:G, [C} G:A, and [D} G:T.

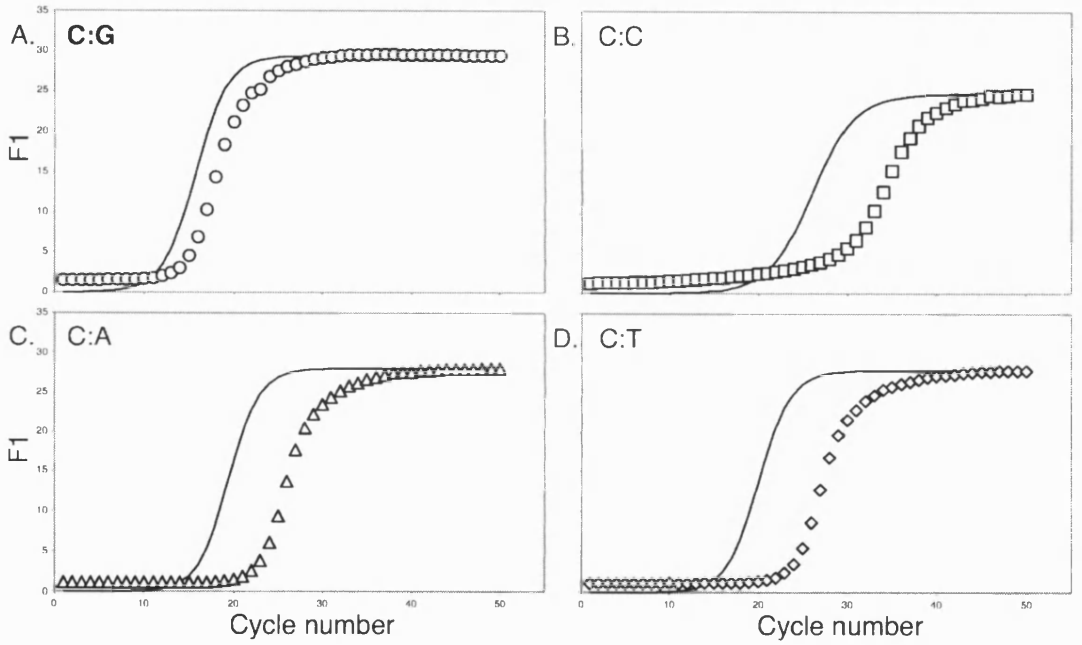


Figure 6.9: Correlation of real time substrate accumulation (data points) with that predicted by the sigmoid function (solid line). Graphs display base combinations (template: primer) [A] C:G, [B] C:C, [C] C:A, and [D] C:T.

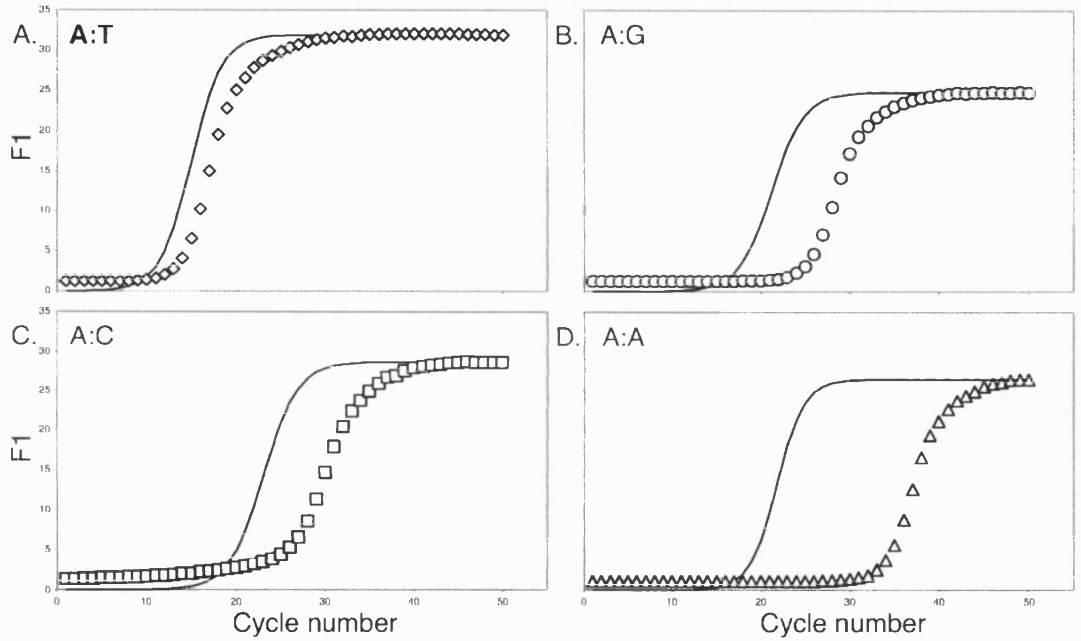


Figure 6.10: Correlation of real time substrate accumulation (data points) with that predicted by the sigmoid function (solid line). Graphs display base combinations (template: primer) [A] A:T, [B] A:G, [C] A:C, and [D] A:A.

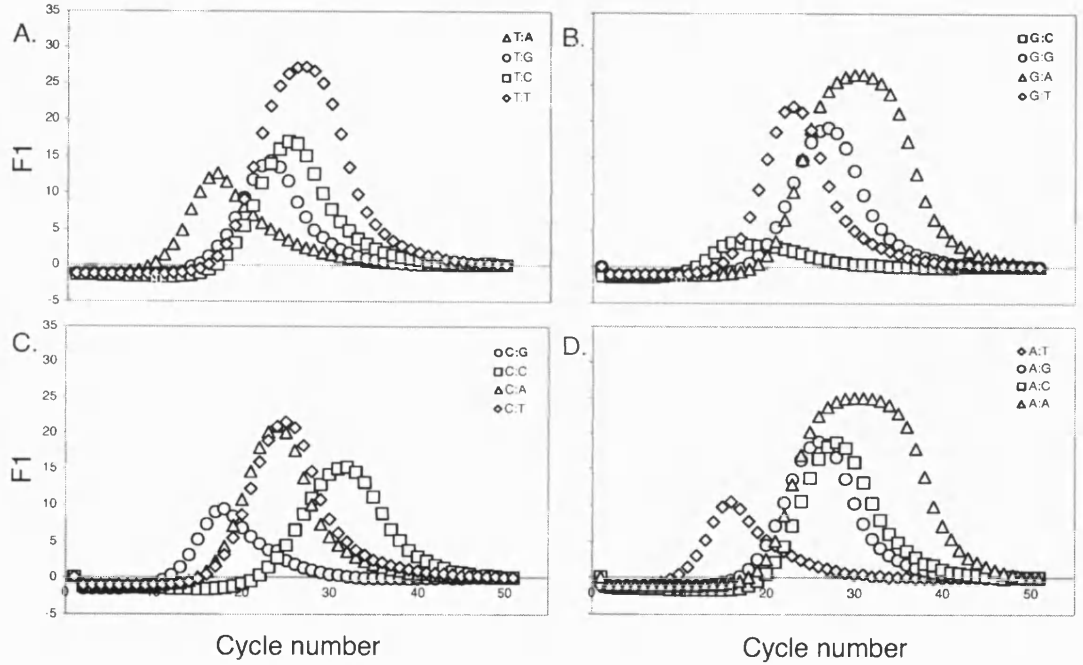


Fig 6.11: Plots showing the difference between expected fluorescence ($F1$) values, as predicted by the kinetic model, and actual fluorescence values according to the LightCycler™ instrument for all SNP combinations.

	Denaturation		*Annealing*		*Extension*		
Set	°C	s	°C	s	°C	s	Cycles
A	95	1	63	5	72	6	60
B	95	1	63	5	72	7	60
C	95	0	55	5	72	10	60
D	95	1	62	5	72	7	60

Table 6.2: Deviations of cycling protocols from that described in Section 2.0.2, used for quantification experiment sets where [A] represents amplifications by primer set ICF/ICR, [B] represents amplifications by primer set SC_WT_OP/SC_MUT_OP, [C] represents amplifications by primer set WT_AS_C/WT_OP_98 and [D] represents amplifications by primer set C282Y_WT_OP/C282Y_MUT_OP.

Reaction	*Sample*	*A: total DNA, ng*	*B: total DNA, ng*	*C: total DNA, ng*	*D: total DNA, ng*
1	STD	20 .0	30.0	30.0	30.0
2	STD	2.0	3.0	3.0	3.0
3	STD	0.2	0.3	0.3	0.3
4	STD	0.02	0.03	0.03	0.03
5	STD	0.002	0.003	0.003	0.003
6	NEG	0.0	0.0	0.0	0.0
7	UKN	10.0	15.0	1.5	15.0
8	UKN	1.0	1.5	0.15	1.5
9	UKN	0.1	0.15	0.015	0.15

Table 6.3: Initial DNA concentrations used for quantification experiment sets where [A] represents amplifications by primer set ICF/ICR, [B] represents amplifications by primer set SC_WT_OP/SC_MUT_OP, [C] represents amplifications by primer set WT_AS_C/WT_OP_98 and [D] represents amplifications by primer set C282Y_WT_OP/C282Y_MUT_OP.

Data points showing the cycle dependency of rate v for the dilution series of reaction sets A to D, were collected using the method described in Section 6.1.2. The kinetic model (equation 6.7) was used in non-linear regression analysis to fit real data from the qPCR tests (Figure 6.12 A-D inclusive) to estimate informative kinetic parameters for each DNA concentration (Table 6.4). The position of the peak along the x-axis differed according to starting DNA concentration, shifting to the right as DNA concentration was decreased. The apex of the curve was used as an independent measure of C_T in the PCR test. The C_T from the known standard dilution series was plotted vs. log known concentration to produce a standard curve; Figure 6.13 A-D).

An identical set of experiments was repeated and analysed using LightCycler™ Data Analysis software (Version 3.0). The 'Fit Points Method' of quantification was used to define the cycle threshold or crossing point (Figure 1.18A, pg. 59). In the Fit Points method, the user manually sets a baseline above background noise. Points are selected from the log-linear portion of the curve, and a straight line of $y = mx + c$ (where y is the C_T value, m is the gradient of the trend line, c is the intercept on the y-axis, and x is $\log_{10}$ [sample]) is drawn through the points. The crossing point is the point at which an extrapolation of the linear curve crossed with the manually set threshold. Crossing point data is plotted vs. $\log_{10}$ sample concentration to produce a standard curve. Actual C_T data from the kinetic model and LightCycler™ analyses is displayed in Table 6.5 (Set A), Table 6.6 (Set B), Table 6.7 (Set C) and Table 6.8 (Set D).

For the kinetic model and LightCycler™ analyses, the linear trend line of the standard curve provides a function, which can be solved for x. Therefore, the starting concentration of unknown samples was estimated (Table 6.9).

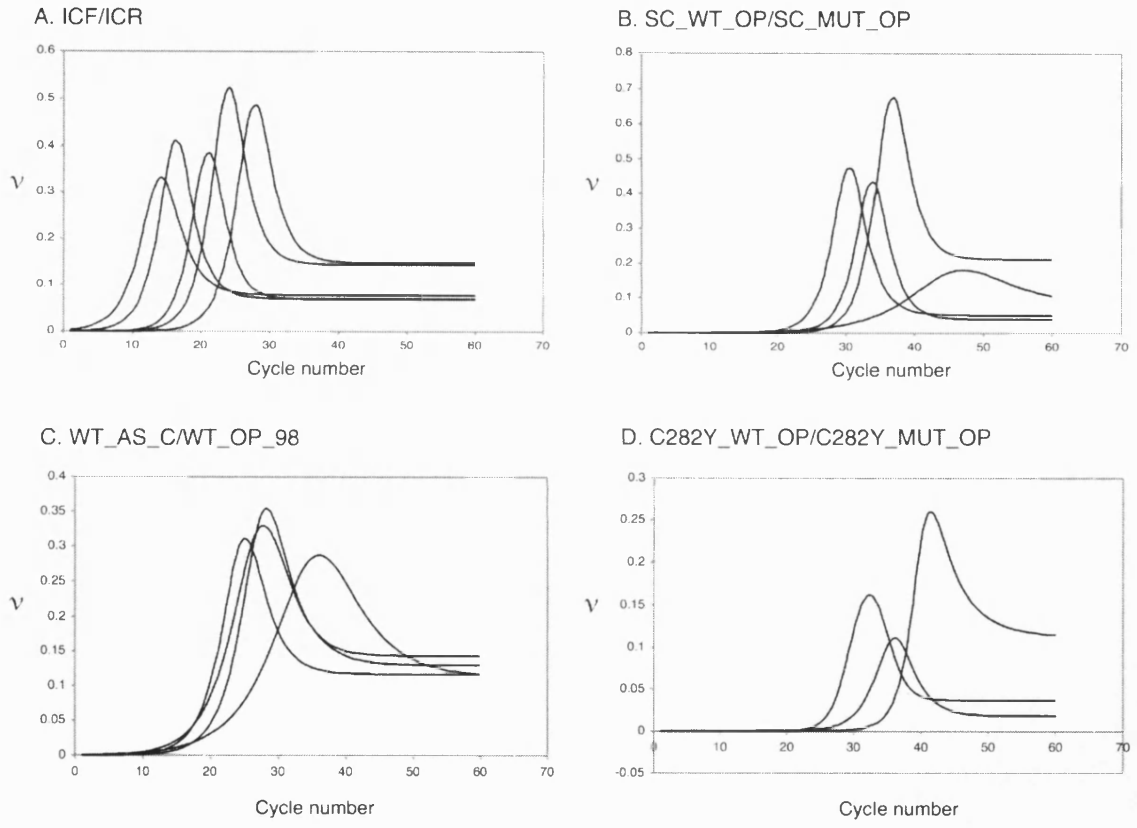


Figure 6.12: Kinetic model fit from real-time continuous monitoring data from amplifications directed by [A] Primer set ICF/ICR [B] Primer set SC_WT_OP/SC_MUT_OP; [C] Primer set PM_AS_C/WT_CP_98; [D] Primer set C282Y_WT_OP/C282Y_MUT_OP. Raw data regarding the initial template concentration associated with each peak is displayed in Tables 6.5 to 6.8.

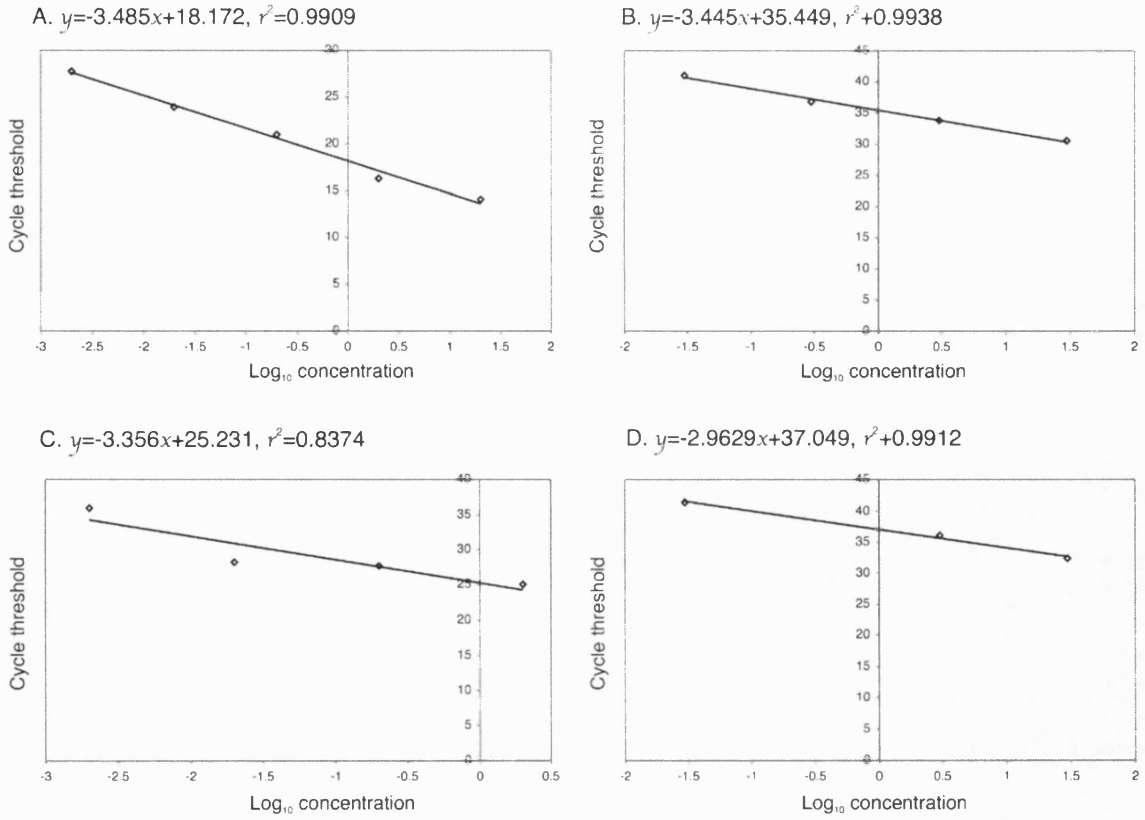


Figure 6.13: External standard peak thresholds (◊) are fitted with a linear trend line to allow estimation of 'unknown' sample concentrations. Equation of trend lines and correlation coefficients are located above the relevant graph from amplifications directed by [A] Primer set ICF/ICR [B] Primer set SC_WT_OP/SC_MUT_OP; [C] Primer set PM_AS_C/WT_CP_98; [D] Primer set C282Y_WT_OP/C282Y_MUT_OP.

Set	*[Sample], ng*	V_{max}	K_M	C	K_i
A	20	610	293	-0.627	3.50E-04
	2	843	852	-0.899	8.21E-04
	0.2	1.28	186	-1.00	138
	0.02	256	2463	-0.986	4.16E-02
	0.002	801	8922	-0.986	1.33E-02
B	30	2700	2650	-0.996	0.008
	3	3940	8410	-0.998	0.014
	0.3	2790	19400	-0.999	0.059
	0.03	1540	44400	-1.000	0.564
	0.003	10.5	119000	-1	61900
C	20	2.96	123	-1	473
	2	7610	814	-1	0
	0.2	7210	2220	-0.996	0
	0.02	7790	1600	-0.995	0
	0.002	8900	5430	-1	0
D	30	7.81	2640	-1	1410
	3	5.94	11100	-1	14300
	0.3	144	25700	-1	25.2
	0.03	4.13	15500	-1	65500
	0.003	0.671	781000	-1	1400000

Table 6.4: Michaelis-Menten parameters estimated by the kinetic model (equation 6.9). Data is displayed for amplifications directed by [A] Primer set ICF/ICR [B] Primer set SC_WT_OP/SC_MUT_OP; [C] Primer set PM_AS_C/WT_CP_98; [D] Primer set C282Y_WT_OP/C282Y_MUT_OP.

Table 6.5: C_T data from amplification set A, predicted by kinetic model vs. LightCycler™ Fit Points Method

Sample	*[Sample], ng*	*Log_{10} [Sample]*	*Kinetic model C_T value*	*LightCycler™ C_T value*
STD 1	20	1.30	14.11	5.66
STD 2	2	0.30	16.29	8.87
STD 3	0.2	-0.70	21.00	11.22
STD 4	0.02	-1.70	23.92	16.25
STD 5	0.002	-2.70	27.72	18.13
NTC	-	-	-	-
UKN 1	10	1.00	14.68	6.64
UKN 2	1	0.00	17.01	9.92
UKN 3	0.1	-1.00	21.46	13.58

Table 6.6: C_T data from amplification set B, predicted by kinetic model vs. LightCycler™ Fit Points Method

Sample	*[Sample], ng*	*Log_{10} [Sample]*	*Kinetic model C_T value*	*LightCycler™ C_T value*
STD 1	30	1.48	30.52	22.82
STD 2	3	0.48	33.81	26.29
STD 3	0.3	-0.52	36.76	29.68
STD 4	0.03	-1.52	41.02	33.57
STD 5	0.003	-2.52	47.11	51.22
NTC	-	-	-	-
UKN 1	15	1.18	32.12	23.77
UKN 2	1.5	0.18	35.45	27.64
UKN 3	0.15	-0.82	38.53	31.14

Table 6.7: C_T data from amplification set C, predicted by kinetic model vs. LightCycler™ Fit Points Method.

Sample	*[Sample], ng*	*Log10 [Sample]*	*Kinetic model C_T value*	*LightCycler™ C_T value*
STD 1	20	1.30	59.63	20.7
STD 2	2	0.30	25.01	22.28
STD 3	0.2	-0.70	27.74	25.09
STD 4	0.02	-1.70	28.24	25.8
STD 5	0.002	-2.70	36.03	32.23
NTC	-	-	45.74	40.19
UKN 1	1	0.00	25.87	23.62
UKN 2	0.1	-1.00	28.92	26.37
UKN 3	0.01	-2.00	29.4	27.26

Table 6.8: C_T data from amplification set D, predicted by kinetic model vs. LightCycler™ Fit Points Method.

Sample	*[Sample], ng*	*Log10 [Sample]*	*Kinetic model C_T value*	*LightCycler™ C_T value*
STD 1	30	1.48	32.35	26.26
STD 2	3	0.48	36.12	29.71
STD 3	0.3	-0.52	-	33.49
STD 4	0.03	-1.52	41.4	38.75
STD 5	0.003	-2.52	60	37.42
NTC	-	-	-	-
UKN 1	15	1.18	1	28.09
UKN 2	1.5	0.18	37.92	31.01
UKN 3	0.15	-0.82	40.75	34.51

Reaction Set	*Actual [sample], ng*	*Kinetic model Calculated log_{10}[sample]*	*Kinetic model Calculated [sample], ng*	*LightCycler™ calculated [sample], ng*
A	10	1.00	10.05	9.05
	1	0.33	2.15	0.91
	0.1	-0.94	0.11	0.70
B	15	0.97	9.25	15
	1.5	-2.9E-04	1.00	1.27
	0.15	-0.89	0.13	0.13
C	1	-0.19	0.65	0.55
	0.1	-1.10	0.079	0.15
	0.01	-1.24	0.057	0.0046
D	1.5	-	-	8.65
	0.15	-0.29	0.51	1.70
	0.015	-1.25	0.056	0.24

Table 6.9: Calculated concentrations of 'unknown' DNA samples using kinetic model vs. LightCycler™ Fit Points Method. Data is displayed for amplifications directed by [A] Primer set ICF/ICR [B] Primer set SC_WT_OP/SC_MUT_OP; [C] Primer set PM_AS_C/WT_CP_98; [D] Primer set C282Y_WT_OP/C282Y_MUT_OP.

6.2 Discussion

6.2.1 *Development of Kinetic Model of PCR*

Figure 6.2 A shows a typical plot of fluorescence $\Delta F1$ during PCR. The observed transitions in fluorescence closely follow the temperature profile because of the strong dependency of fluorescence on temperature (Wittwer *et al.,* 1997). As the sample was heated, fluorescence was high until denaturation occurs (apparent as a sharp drop in fluorescence). As the sample was cooled from denaturation to annealing temperature, fluorescence increased rapidly, reflecting product–to-product annealing (Figure 6.2 B). Fluorescence also increased during extension whilst the temperature was held within the optimal limits for *Taq* DNA polymerase (Figure 6.2 C). This increase was directly attributable to polymerisation and the accumulation of double stranded DNA, and was measured to represent the rate, v, of each individual cycle during the PCR.

The dependency of reaction rate, v, on substrate concentration $[S]$ for all PCR tests showed a strong departure from the hyperbolic relationship predicted by the Michaelis-Menten equation (Figure 6.1 A). In fact, the reaction rates generated from real time data closely followed the general curve predicted for high substrate inhibition (Figure 6.1B; Equation 6.5).

To fit these data using equation 6.5 directly would not provide a good approximation, since PCR is a non-classical enzyme system. In PCR the product formed in cycle $n-1$ is the substrate that is available for cycle n, and therefore cannot be represented simply using a linear scale. Under non-limiting conditions this should result in an exponential increase in substrate concentration over successive cycles described in the equation,

$$S_n = 2^n S_0, \tag{6.9}$$

where S_0 is the initial template (substrate) concentration, and S_n is the substrate concentration after n successive cycles. It is well understood, however, that in PCR

this model only holds for early cycles due to the changing dynamics in the plateau phase (Schnell and Mendoza, 1997A).

Consequently, a simplified sigmoidal function was applied to describe the accumulation of substrate during PCR,

$$S = \frac{1}{(1 + a * \exp^{(-f*n)})} \tag{6.8}$$

where a and f are descriptive parameters of the sigmoid function at cycle n (Figure 6.15; Seggern, 1993). This function arises in many dynamical systems, describing simple exponential growth dynamics with a linear limiting control. Such systems are often called logistic growth. A similar sigmoidal function was recently reported as providing a good fit for the whole kinetic process of real time PCR (Liu and Saint, 2002).

The model was further refined by the inclusion of a compensation factor C. Equation 6.5 demands that at zero substrate concentration the rate of the reaction is zero. Since a PCR contains an initial substrate (template) concentration at cycle number n = 1, this compensation factor allowed the criteria of the equation to be satisfied, such that an artificial cycle 0 is included. The overall equation becomes,

$$v = \frac{V_{\max}}{\left[1 + \left(\frac{K_M}{(S+C)}\right) + \left(\frac{(S+C)}{K_i}\right)\right]} \tag{6.7}$$

where S is represented by the sigmoidal function of equation 6.8. This role of the compensation factor was confirmed by the consistent prediction of C as approximately 1 arbitrary unit (Tables 6.1 and 6.4).

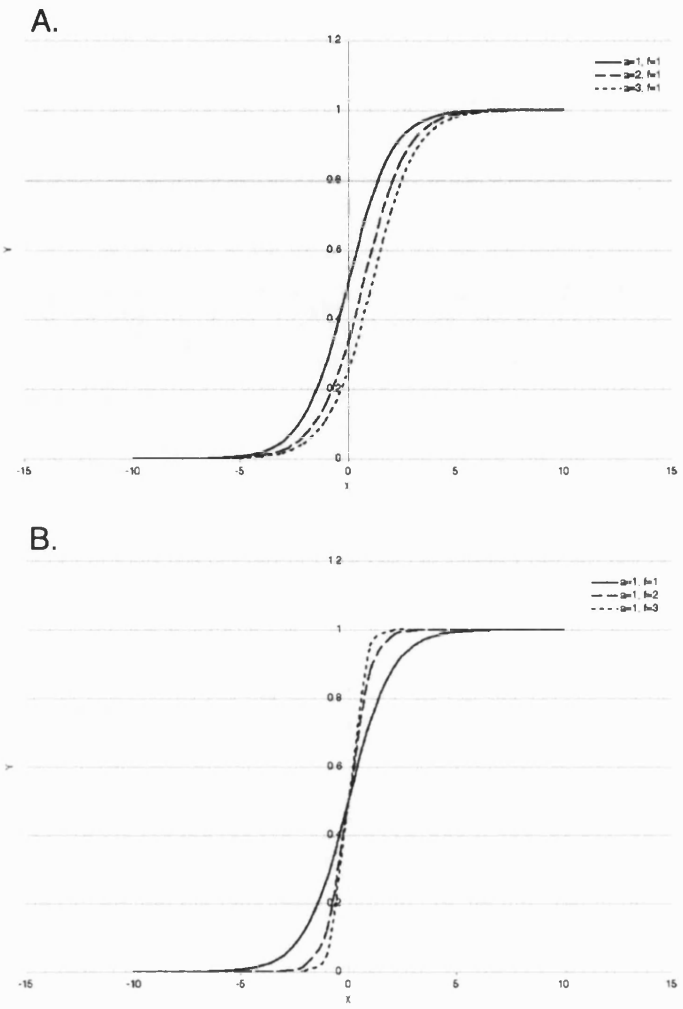


Figure 6.15: The effects of varying parameters [A] a and [B] f on the general shape of a sigmoid curve. The sigmoid function was solved using the numbers shown in the legend, and plotted alongside each other for comparison.

The function $(S+C)$ approximated well to real-time data in these experiments (Figures 6.7-6.10 A-D inclusive), particularly in correctly primed PCR tests (Figure 6.11). For general applications such as quantitative PCR, this model is targeted for use in correctly primed PCR tests, and therefore the greater non-conformity of the model in the presence of a mismatch would not affect the utility of the test. The difference is likely due to other intrinsic factors that occur when a mismatch is present. These are discussed in the following section.

6.2.2 *Determination of kinetic constants in allele-specific PCR*

The synthesis of oligonucleotides that differ only at the 3'-terminal base, coupled with templates with a variant nucleotide at a defined position, provided a system to evaluate the effects of mismatches in PCR. The ability of *Taq* DNA polymerase to bind and extend terminal mismatches was dependent upon the identity of the mispair, inferred by the differing kinetic parameters estimated (Table 6.1).

It was assumed that $V_{\max}$ represented the maximal rate of incorporation of SYBR into an extending DNA complex under non-limiting conditions. This provides an arbitrary measure of maximal polymerisation or catalytic activity. Although $V_{\max}$ is not a fundamental property of an enzyme as it is dependent on enzyme concentration (Cornish-Bowden, 1979), reactions were validly compared using this parameter since the same total enzyme concentration, E_T, was used in each PCR test.

For amplifications from template with a variant nucleotide G, $V_{\max}$ was highest in the presence of a matched primer 3' terminus. For all other templates at least one of the mismatched primers gave a higher $V_{\max}$ value than the match. In PCR tests involving the transversion mispairs (template:primer) G:A, A:A and T:T, $V_{\max}$ was lower than that found for a match. The transition mispairs C:A, A:C, T:G and transversion mispairs A:G and T:C exhibited equal or higher $V_{\max}$ than a matched combination.

The latter observation is not surprising when one considers the consequence of mismatched primer extension in PCR. Certain nucleotide configurations vary in their affinity to bind to the enzyme and extend. However, once extension from a mismatched primer occurs, the resultant product and the complement synthesised in subsequent cycles are fully matched with both primers (Kwok *et al.,* 1990). The nature of PCR means that a single copy of matched template can be amplified to plateau levels in just a few cycles. This causes failure in many end-point genotyping tests by impairing discrimination. Therefore, $V_{\max}$ alone is insufficient to ascertain the true effects of a mispair in PCR.

The more dominant parameter for characterising enzymatic discrimination is the Michaelis constant K_M, defined as the substrate concentration at which the reaction rate is half its maximal velocity (Cornish-Bowden, 1979). The K_M is an intrinsic property of an enzyme reflecting the binding constant for forming the enzyme substrate complex, as well as the catalytic constant. In PCR, this embodies template affinity and the thermodynamic environment of the reaction (Schnell and Mendoza, 1997B). The transversion mispairs G:A, A:A, and T:T exhibited a K_M value between 100- to 1000-fold higher than a corresponding match. In all other cases, K_M for a mismatch was 8- to 33- fold higher than the equivalent match. These data suggest that a matched conformation binds with higher affinity than a mismatch to the active site of the enzyme.

This is discordant with the view of Huang *et al.* (1992) who found that *Taq* polymerase binds with equal affinities to all base configurations. This is likely due to the reaction conditions employed where a single species of dNTP was used as the variable substrate for each template. Single-turnover kinetics is unable to reflect a PCR test in which all dNTPs are present at equimolar concentration and in large excess. In the absence of competition, similar binding affinities might be expected. Some groups were similarly conflicting (D'yachenko *et al.,* 1994; Kuchta *et al.,* 1987), whilst others were in agreement (Day *et al.,* 1999; Mendelman *et al.,* 1990; Petruska *et al.,* 1988; Wong *et al.,* 1991). The diversity of reaction conditions employed, enzyme source and DNA sequences may have caused this disparity.

Importantly, the work presented during this chapter represents the kinetic characterisation of *Taq* polymerase in a true rapid cycle PCR environment overcoming the limitations of previous findings.

The kinetic parameters displayed in Table 6.1 were used to define a value for the fundamental 'specificity' constant, k_{cat}/K_M, where k_{cat} is the catalytic constant or turnover number. This specificity constant determines the ratio of reaction rates for enzyme acting on two competing substrates, when they are mixed together at equimolar concentrations. Given that it is the ratio of specificity constants that determines the ratio of rates of the competing reactions *a* and *b*, V_{max}/K_M was evaluated since $[E_T]$ was identical throughout these tests (Cornish-Bowden, 1979):

$$\frac{\left(V^A_{max}/K^A_M\right)a}{\left(V^B_{max}/K^B_M\right)b} = \frac{\left(\left(V^A_{max}/E^A_T\right)/K^A_M\right)a}{\left(\left(V^B_{max}/E^B_T\right)/K^B_M\right)b} = \frac{\left(k^A_{cat}/K^A_M\right)a}{\left(k^B_{cat}/K^B_M\right)b}$$

$$\frac{\left(V^A_{max}/K^A_M\right)a}{\left(V^B_{max}/K^B_M\right)b} = \frac{\left(\left(V^A_{max}/E^A_T\right)/K^A_M\right)a}{\left(\left(V^B_{max}/E^B_T\right)/K^B_M\right)b} = \frac{\left(k^A_{cat}/K^A_M\right)a}{\left(k^B_{cat}/K^B_M\right)b} \qquad (\text{if } E^a_T = E^b_T) \qquad (6.10)$$

The comparison of these values between matched- and mismatch-primed PCR expresses the enzyme's ability to discriminate in favour of a particular base configuration in the presence of others (Cornish-Bowden, 1979), and consequently the relative extension efficiencies (Huang *et al.*, 1992). In agreement with work by Kwok *et al.* (1990) it was found that a match was the more specific substrate for *Taq* DNA polymerase. The relative specificity of mismatches varied according to each arrangement, but in general, purine:purine mispairs were harder to extend than other combinations. This system unequivocally demonstrates that for *Taq* DNA polymerase during rapid cycle PCR, the matched base configuration would be the preferred substrate in the presence of all other mismatches.

The model also provided estimate values for the inhibition constant K_i, which provides a quantitative measure of inhibitor potency (Wharton and Eisenthal, 1981). Values for K_i are usually derived from an inhibition study of the effects of a substance on kinetics or binding of a substrate or effectors. Such studies are usually performed in the presence of the true substrate (for kinetics) or ligand (for

binding) AND the inhibitor. Therefore the K_i value in this model represents a substrate inhibition constant since the substrate only is present. This could be interpreted as the ability of the mismatched primer to form a dead-end complex by binding to the active site of the enzyme without efficient extension. In addition, it may represent the ability of the primers to bind dystopically to a particular template, such that replication rate is depressed. In both cases, one might expect K_M for a mismatched primer to be higher than that of the true primer. This is supported by data in which the three transversion mispairs, G:A, A:A, and T:T, exhibited a significantly higher K_i value than all other mispairs. Consistently, K_i was lower for a match. By considering the other parameter values associated with each configuration, it is proposed here that these particular mispairs bind to the enzyme with low affinity and extend with low efficiency, forming a stable dead-end complex that inhibits progression of the reaction.

Due to the complex interactions of components within a PCR (Linz *et al.,* 1990), and the thermodynamics of DNA hybridisation (Petruska *et al.,* 1988), variations in component concentrations or annealing temperature during the test will also alter the resulting kinetic constants, not just the inhibition constant. It is therefore important to only compare kinetic parameters obtained under identical reaction conditions.

6.2.3 High substrate inhibition

Factors that have been attributed to attenuation of PCR include depletion of substrate (dNTPs or primers), thermal inactivation or limiting concentration of DNA polymerase, inhibition of enzyme activity by increasing pyrophosphate production, reannealing of amplicon at concentrations above 10^{-8} M, reduction in the denaturation efficiency per cycle, destruction of product due to enzyme 5'-3' exonuclease activity (Kainz, 2000), product to product reannealing (Wittwer *et al.,* 1997) or the chelation of critical metal ions by the substrate depriving the enzyme of a cofactor (Wharton and Eisenthal, 1981). For each theory, the characteristics of high substrate inhibition would be apparent.

Previous investigations have shown that by including increasing amounts of "random" DNA into the PCR test from the beginning amplification could be inhibited (Kainz, 2000; Dimitrov and Apostolova, 1996). It was suggested that the accumulation of product during later cycles is more likely to inhibit the enzyme before the more trivial factors such as exhausting primers, metal ions or dNTPs, which are added in huge molar excess. This, in conjunction with the excellent fit of the kinetic model to real data provides compounding evidence that high substrate inhibition does indeed play a predominant role in curtailing amplification during the final cycles of a PCR.

However, one might have expected the K_i for a match to be greater than or equal to a mismatch if we were proposing high substrate inhibition as the sole cause of the PCR plateau phase. These data show that in the case of a mismatch, other factors come into play, which further add to the magnitude of the K_i value. This is supported by Figures 6.7-6.10 A-D inclusive, in which the model's prediction of substrate accumulation was more accurate for a match- than mismatch-primed PCR test. Figure 6.11 highlights this trend further. So although high substrate inhibition is the likely cause of the plateau phase in conventional PCR, it seems likely that the actual presence of a mismatch has an effect on the intrinsic ability of the enzyme to function effectively, displaying a direct inhibitory effect on the PCR.

6.2.4 PCR quantification using external standards

Absolute and relative quantification approaches can compare amplification of a target nucleic acid in samples of unknown DNA concentration, against a standard curve consisting of serial dilutions with known concentrations of the same target (homologous standards) or a reference gene. The standards are generally run in different reaction vessels, but within the same PCR run (external standards). For absolute quantification, target values are usually expressed as an absolute value (copy number or nanograms). In the case of relative quantification using a standard curve, the quantity is expressed relative to a reference/housekeeping gene, or calibrator. Instead of an absolute value, for all experimental samples the target quantity is determined from the standard curve and divided by the target quantity of

the calibrator. Thus the calibrator becomes the 1x sample, and is used to "normalise" all other quantities, which are expressed as an n-fold difference relative to the calibrator.

For either method to be accurate and reliable, a single number is required that consistently identifies a single defined position on the curve, the C_T value. The kinetic model was used to define the cycle threshold based on kinetic parameters. The peak of each kinetic curve (Figure 6.12 A-D) represents the transition point where the substrate concentration becomes limiting (Figure 6.16). It is assumed that since this is dependent on enzyme concentration and sequence only (not withstanding experimental error), that this point is consistent regardless of template concentration. The good correlation between initial template concentration and the calculated threshold confirmed the accuracy of the model (Tables 6.5-6.8 inclusive).

In these experiments, the cycle number at which this threshold occurred increased in accordance with starting template concentration. A standard curve was constructed from these data to produce a linear trend line (Figure 6.13 A-D) from which the initial DNA concentrations of unknown samples were calculated (Table 6.9). The kinetic model was directly compared with the Fit Points Method used in the LightCycler™ (Tables 6.5-6.8 inclusive). The software used by the LightCycler™ did not allow the Fit Points Method to be used with continuous monitoring. Therefore, deviation not only represents pipetting error and other run-to-run variations, but also reflects differences in the way in which real time data was obtained.

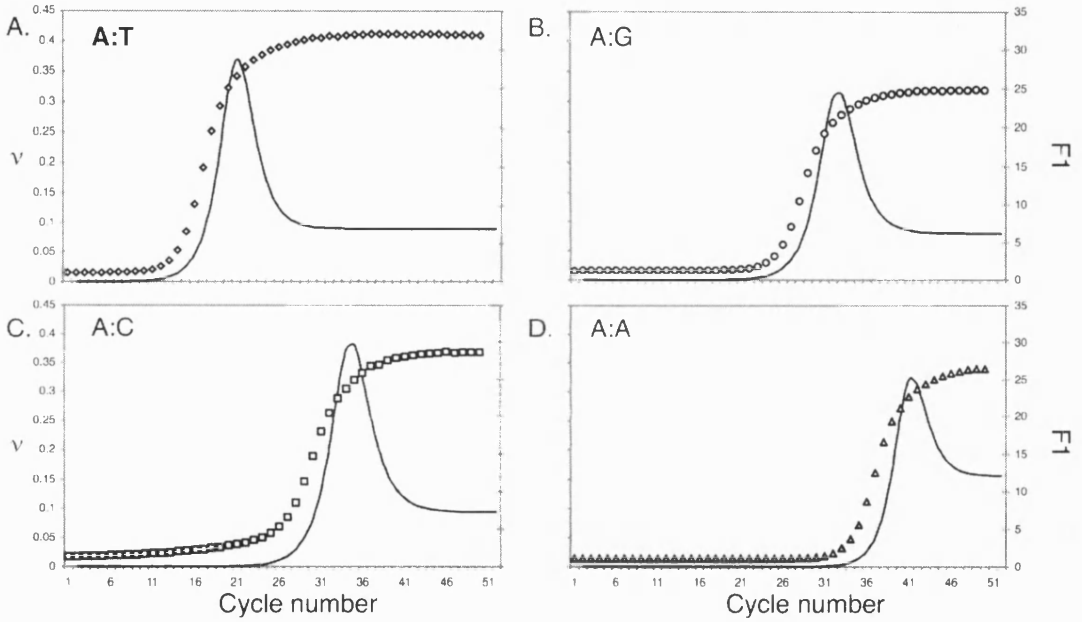


Figure 6.16: Real-time PCR data (points) overlaid by the kinetic model curve for the corresponding reaction. Figure demonstrates that the point used to derive the C_T value is consistent for each reaction.

To summarise, a kinetic model of PCR was developed in which the PCR progress curve was viewed as a series of independent reactions where initial rates were accurately measured for each cycle. The variation of reaction rates per cycle was shown to closely follow the general equation for high substrate inhibition, which formed the foundation of the kinetic model, demonstrated here to have high correlation to real data. The general trends presented for matched and mismatched amplifications are directly applicable to the design of SNP genotyping experiments, providing a measure of the likely success and optimisation necessary.

This approach has also been validated for application in rapid qPCR methods in which accurate measurement of quantitative parameters is essential. The method described here can be, in principle, applied to probe detection formats as well as the SYBR Green I sequence non-specific detection. For example, if TaqMan® probes were used, an increase in reaction rate would be observed at the annealing/extension phase of the reaction as the 5' nuclease activity is most effective. As the amount of cleavage is directly proportional to target DNA concentration, one might expect a similar curve of initial rate vs. cycle number to be formed as observed here. These data would be amenable to the kinetic model described here. This concept has not been validated, and represents an area where further research can be conducted.

This work forms the basis of Patent No, GB02_21295.9 "Improvements in or relating to nucleic acid amplification reactions" and was published in Biochemical and Biophysical Research Communications, 2002 (Appendix VI).

Chapter Seven

General Discussion

SNP genotyping methodologies are under pressure to deliver technologies that work rapidly, are robust, have the potential for automation, and a reduced requirement for controls. Combined, these attributes will offer a decreased need to re-run assays, leading to cost efficiency, the major pressure for assay development. The successful assays of the future will need to reduce costs by utilising less expensive reagents and technology, but also through improving robustness and automation. The aim of this thesis was to develop an assay that could fulfil these key requirements. Ideally, the test should be suitable for use in a clinical environment for routine testing, involving few steps once optimised to avoid the potential for contamination and sample mix-ups. Understandably, low cost and high speed is important here.

PCR, a widely used technique in life science research, provided the ideal platform to satisfy these criteria. PCR is utilised in almost all genotyping techniques to achieve the required sensitivity and specificity, and was used as the core technology in the work of this thesis. Of particular relevance is the adaptation of PCR itself called allele-specific PCR. Although this methodology is robust, accurate, and the method of choice for routine screening of sickle cell anaemia for example (Old *et al.,* 2000), it has been limited previously by lengthy optimisation strategies, the requirement of two-tubes to genotype a bi-allelic SNP, the necessity of internal control primers, and labour-intensive post-PCR processing using gel-based techniques.

Consequently, a single tube adaptation of allele-specific PCR was developed during this work called bi-directional PCR (presented in Chapter Three), in which allele-specific amplicons are separated according to size on an agarose gel. The assay was validated by high accuracy in blinded trials for the detection of the A→T

mutation in the β-globin gene linked to sickle cell anaemia. This test would likely be valuable in laboratories where contemporary instrumentation is unavailable, and throughput requirements are moderate. Current drawbacks of the assay are low-speed and the requirement for multiple-steps.

Bi-directional PCR was developed further, by introducing homogeneity to the test using a universal dsDNA specific binding dye, SYBR® Green I, to monitor specific product accumulation the PCR reaction mixture. This enabled real time detection of amplification and identification of allele-specific products using melt curve analysis. Figure 7.1 demonstrates how these attributes interact to address the requirements of the perfect assay (Figure 1.4, pg. 16). The fluorescent bi-directional PCR test was validated by high accuracy in blinded trials for the detection of the A→T mutation in the β-globin gene linked to sickle cell anaemia, and the two most common mutations linked to increased susceptibility to hereditary haemochromatosis.

The fluorescent bi-directional PCR assay could be improved for more widespread application by introducing automation with the use of robotics systems for reaction set-up, and computer software for primer design. With improved temperature control and measurement in PCR instruments, a slower more-controlled melt curve analysis could allow products differing in T_m by less than 2°C to be effectively discriminated. The test could be applied to TrueSNP™ primers, a phosphodiester primer with a LNA monomer at the 3' end. This product was designed solely to improve specificity in allele-specific PCR tests, validated by brief experiments (Appendix V). The cost of TrueSNP™ primers is approximately double the cost of a conventional primer, but approximately one-fifth of the cost of a single real-time PCR probe. Since the true cost of a genotyping test is in the optimisation strategy as well as the test itself, additional cost benefits could be seen in optimisation, where the speed and likelihood of establishing discriminatory conditions could be increased.

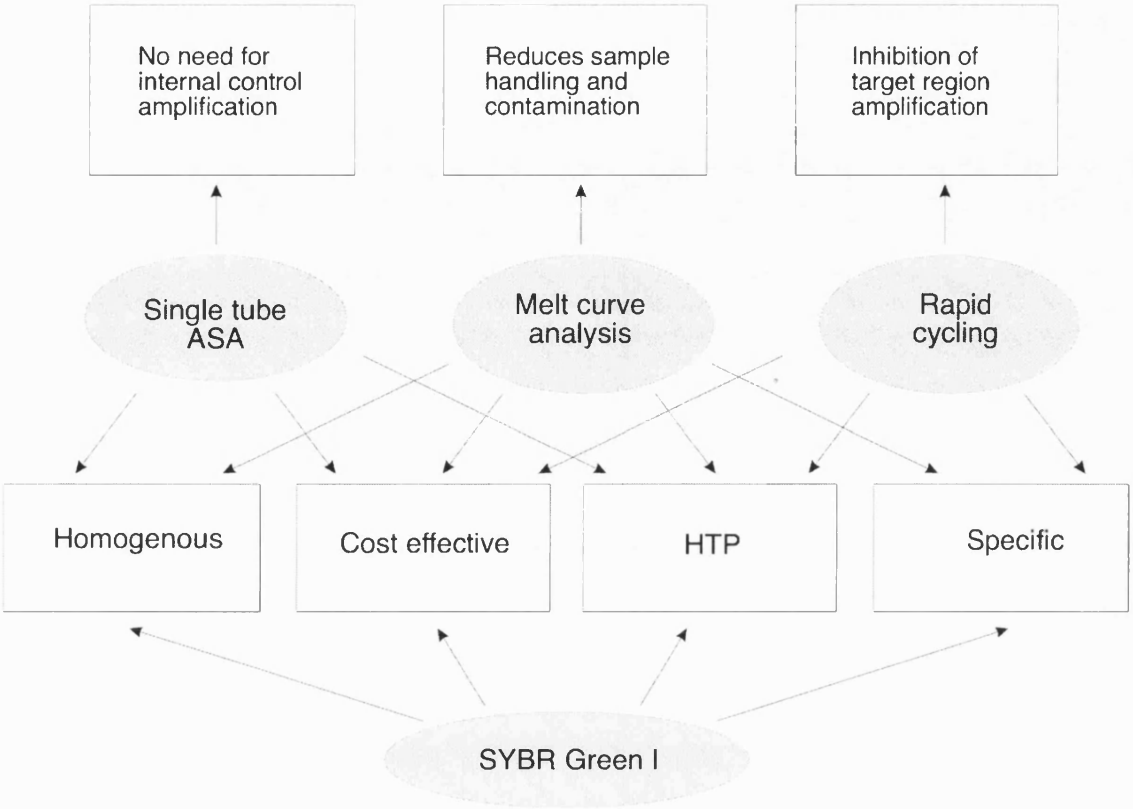


Figure 7.1: A schematic representation of the four key components of the fluorescent bi-directional PCR assay (shown in grey), and their favourable interactions. HTP represents high throughput.

However, where sequence specificity is of paramount importance, real-time PCR probes are the current tools of choice. During this work, the bi-directional PCR was adapted to use TaqMan® probes for allele-specific amplicon detection, as opposed to a universal DNA binding dye. This assay is semi-sequence specific since a specific amplicon is generated following allele-specific amplification influenced by the enzyme and PCR conditions, not the sequence-specific probe. Therefore, of the three assays developed here, this adaptation conferred the fewest advantages in terms of speed and specificity, and incorporated additional costs and design considerations using real-time PCR probes.

The importance of SNP detection assays, such as those developed during this thesis has been realised recently. Following the sequencing of the human genomic project, the acceptance of genetic testing has escalated because it has been recognised that SNPs are linked not only to monogenic disorders of which 1600 are known, but also to increased susceptibility to common multifactorial disorders such as cancer and heart disease, and the prediction of drug efficacy and toxicity. Therefore, genetics and SNP testing strategies will need to be introduced into drug development.

Pharmacogenomic studies are already underway to look at the genes involved in drug metabolising enzymes or in routinely characterising the genetic variation in potential drug targets. Areas such as oncology would significantly benefit from a prediction of therapeutic affect based on genetic tests due to the multiple side effects and varying responses exhibited by patients administered these drugs. Where a typical choice of treatment exists, it would be beneficial to examine a patient's response to that drug, quickly, prior to administration. The *trial-and-error* approaches used today in which sequential drugs are administered until a favourable response is apparent can be expensive and unacceptable to patients. This approach also increases the risk of cumulative side affects and non-compliance (Sanders, 2002). Oncology represents one of many healthcare areas that would potentially benefit in similar ways.

PCR-based testing has had huge impact on other areas of diagnostic testing, such as the monitoring of gene expression in certain disease states, or viral load or bacterial particle quantification (Chapter Six). The most accurate and reproducible method for performing these analyses is qPCR techniques. Currently, there are a number of instruments available to allow real-time monitoring which has enabled gene quantification based on data collected from the log-linear phase of the PCR. A novel kinetic model developed during this work, was originally used to define useful kinetic parameters (including V_{max}, K_M and K_i) for PCR tests primed by matched and mismatched primers. These data were directly applicable to the design of allele-specific PCR tests. Since the model was based on the internal dynamics of the PCR test itself, it was able to provide an accurate method for PCR quantification using standard curves. The model assumes that PCR becomes limited by high substrate, high substrate inhibition, and good degree of correlation was achieved. The model was also used to determine the cycle threshold in a PCR reaction curve to provide quantitative estimations of known and unknown template concentration. Therefore, this strategy can be used to analyse the PCR test, quantifying specific markers for disease progression. Ultimately, this will contribute to the concept of 'disease management' to optimise the clinical and economical outcome.

In summary, the implementation of genetic testing in routine diagnostics is thought to become a standard and essential part of patient care. The large number of assay formats being produced suggests that soon the goal of a "perfect assay" will be met. Combining the assay with automated nucleic acid extraction systems, automated reaction set-up, and universal internal controls will significantly shorten the turnaround time for generating informative results used for the diagnosis and treatment of patients. In addition, a more thorough understanding of the molecular basis of disease will assist drug discovery, and ultimately allow patient care based on an optimised genetic profile of the individual. It still stands that this can only be achieved with suitable assays for the detection of genetic variation, and it is expected that PCR-based methodologies will be the core of the assay.

References

Allen, D.R., Browse, N.L., Rutt, D.L., Butler, L., Fletcher, C. (1988) The effect of cigarette smoke, nicotine, and carbon monoxide on the permeability of the arterial wall. *Journal of Vascular Surgery*, **7** 139-152

Akey, J.M., Sosnoski, D., Parra, E., Dios, S., Hiester, K., Su, B., Bonilla, C., Jin, L., Shriver, M.D. (2001) Melting curve analysis of SNPs (McSNP®): A gel-free and inexpensive approach for SNP genotyping. *BioTechniques*, **30** (2) 358-367

Andrews, N.C. (1999) Disorders of iron metabolism. *New England Journal of Medicine*, **341** 1986-1995

Ayyadevara, S., Thaden, J.J., Shmookler Reis, R.J. (2000) Discrimination of primer 3'-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction, *Analytical Biochemistry*, **284** 11-18.

Baner, J., Nilsson, M., Mendel-Hartvig, M., Landegren, U. (1998) Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Research*, **26** 5073-5078

Barnes, W.M., Rowlyk, K.R. (2002) Magnesium precipitate hot start method for PCR. *Molecular and Cellular Probes*, **16** 167-171

Baty, D., Kwiatkowski, T.A., Mechan, D., Harris, A., Pippard, M.J. Goudie, D. (1998) Development of a multiplex ARMS test for mutation in the HFE gene associated with hereditary haemochromatosis. *Journal of Clinical Pathology*, **51** 73-74

Becker-Andre, M., Hahlbrock, K. (1989) Absolute mRNA quantification using the polymerase chain reaction: A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Research*, **17** 9437-9446

Bennett, I.C., Gattas, M., The, B.T. (1999) The genetic basis of breast cancer and its clinical implications. *Australian and New Zealand Journal of Surgery*, **69** 95-105

Bertina, R.M., Koeleman, B.P.C., Koster, T., Rosendaal, F.R., Dirven, R.J., de Ronde, H. (1994). Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*, **369** 64-67

Bi, W., Stambrook, P.J. (1997) CCR: a rapid and simple approach for mutation detection. *Nucleic Acids Research*, **25** (14) 2949-2951

Birch, D.E., Kolmodlin, L., Wong, J., Zangenberg, G.A., Zoccoli, M.A. (1996) Simplified hot start PCR. *Nature*, **381** 445-446

Bookchin, R.M., Lew, V.L. (2002) Sickle cell dehydration: mechanisms and intervention. *Current Opinion in Haematology*, **9** 107-110

Bookchin, R.M., Nagel, R.I., Ranney, H.M. (1970) The effect of beta 73 Asn on the interactions of sickling haemoglobins. *Biochimica et Biophysica Acta*, **221** (2) 373-375

Boon, E.M., Ceres, D.M., Drummond, T.G., Hill, M.G., Barton, J.K. (2000) Mutation detection by electrocatalysis at DNA-modified electrodes. *Nature*, **18** 1096-1100

Botstein, D., White, R.L., Skolnick, M., Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32** 314-331

Bottema, C.D.K., Sommer, S.S. (1993) PCR amplification of specific alleles: Rapid detection of known mutations and polymorphisms. *Mutation Research, **288** 93-102*

Breen, G., Harold, D., Ralston, S., Shaw, D., St. Clair, D. (2000) Determining SNP allele frequencies in DNA pools. *BioTechniques*, **28** 464-470

Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Science USA*, **83** 3746-3750

Brightwell, G., Wycherley, R., Waghorn, A. (2002) SNP genotyping using a simple and rapid single-tube modification of ARMS illustrated by analysis of 6 SNPs in a population of males with FRAXA repeat expansions. *Molecular and Cellular Probes*, **16** 297-305

Brookes, A.J. (1999) The essence of SNPs. *Gene*, **234** 177-186

Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, *Journal of Molecular Endocrinology*, **25** 169-193

Carlsson, J. Olsson, S. (2001) Hereditary haemochromatosis and Iron metabolism. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine EJIFCC*, **13** (2)

Chamberlin, M.J. Berg, P. (1963) Comparative properties of DNA, RNA, and hybrid homopolymer pairs. *Federal Proceedings*, **24** 1446-1457

Chang, J.C., Kan, W.K. (1982) A sensitive new prenatal test for sickle-cell anemia. *New England Journal of Medicine*, **307** 30-36

Chehab, F.F., Doherty, M., Cai, S., Wai Kan, Y., Cooper, S., Rubin, E.M. (1987) Detection of sickle cell anaemia and thalassemias. *Nature*, **329** 293-294

Chen, J., Germer, S., Higuchi, R., Berkowitz, G., Godbold, J., Wetmur, J.G. (2002) Kinetic polymerase chain reaction on pooled DNA: A high-throughout, high-efficiency alternative in genetic epidemiological studies. *Cancer Epidemiology, Biomarkers and Prevention*, **11** 131-136

Chen J., Iannone, M.A., Li, M.S., Taylor, J.D. Rivers, P. (2000) A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Research*, **10** 549-557

Chen, X., Kwok, P-Y. (1999) Homogenous genotyping assays for single nucleotide polymorphism with fluorescence resonance energy transfer detection. *Genetic Analysis: Biomolecular Engineering*, **14** 157-163

Chen, X., Levine, L., Kwok, P-Y. (1999) Fluorescence Polarisation in homogenous nucleic acid analysis. *Genome Research*, **9** 492-498

Chen, X., Livak, K., Kwok, P-Y. (1998) A homogenous, ligase-mediated DNA diagnostic test. *Genome Research*, **8** 549-556

Christopherson, C., Sninsky, J., Kwok, S. (1997) The effects of internal primer – template mismatches on RT-PCR: HIV-1 model studies, *Nucleic Acids Research*, **25** 654-658.

Cobb, B.D., and Clarkson, J.M. (1994) A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. *Nucleic Acids Research*, **22** (18) 3801-3805

Compton, J. (1991) Nucleic Acid Sequence-Based Amplification. *Nature*, **350** 91-92

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L., Wallace, R.B. (1983) Detection of sickle cell β-globin by hybridisation with specific oligonucleotides. *Proceedings of the National Academy of Science USA*, **80** 272-282

Cooke G.S., Hill, A.V.S. (2001) Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics*, **2** 967-977

Cornish-Bowden, A. (1979) Fundamentals of Enzyme Kinetics, Butterworth & Co (Publishers) Ltd

Cosa, G., Focsaneanu, K.-S., McLean, J.R.N., McNamee, J.P., Scaiano, J.C. (2001) Photophysical properties of fluorescent DNA-dyes bound to single- and double- stranded DNA in aqueous buffered solution. *Photochemistry and Photobiology*, **73** (6) 585

Crockett, A.O., Wittwer, C.T. (2001) Fluorescein-labelled oligonucleotides for real-time PCR: Using the inherent quenching of deoxyguanosine nucleotides. *Analytical Biochemistry*, **290** 89-97

Davies, S.C., Cronin, E., Gill, M., Greengross, P., Hickman, M., Normand, C. (2000) Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research. *Health Technology Assessment*, **4** (3)

Davignon, J., Gregg, R.E., Sing, C.F. (1988) Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis, **8** 1-21

Day, J.P., Bergstrom, D., Hammer, R.P., Barany, F. (1999) Nucleotide analogs facilitate base conversion with 3' mismatch primers. *Nucleic Acids Research*, **27** 1810-1818

De Silva, D., Wittwer, C.T. (2000) Monitoring hybridisation during polymerase chain reaction. *Journal of Chromatography B*, **741** 3-13

De Silva, D., Reiser, A., Herrmann, M., Tabiti, K., Wittwer, C. (1998) Rapid genotyping and quantification on the Lightcycler™ with Hybridisation Probes. *Biochemica*, **2** 12-15

Didenko, V.V. (2001) DNA Probes using Fluorescence Resonance Energy Transfer (FRET): Designs and Applications. *BioTechniques*, **31** 1106-1121

Dimitrov, D.S., Apostolova, M.A. (1996) The limit of PCR amplification, *Journal of Theoretical Biology*, **178** 425-426

Dubiley, S., Kirillov, E., Mirzabekov, A. (1999) Polymorphism analysis and gene detection by Minisequencing on an array of gel-immobilised primers. *Nucleic Acids Research*, **27** e19

D'yachenko, L.B., Chenchik, A.A., Khaspekov, G.L., Tatarenko, A.O., Bibilashvili, R.S. (1994) Efficiency of DNA synthesis initiated by complementary and mismatched primers. *Molecular Biology*, **28** 654-660

Eisenthal, R. and Danson, M. (2002) Enzyme Assays: A practical approach. Blackwell Publishers Ltd.

Emahazion, T., Feuk, L., Jobs, M., Sawyer, S.L., Fredman, D., St Clair, D., Prince, J.A., Brookes, A.J. (2001) SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends in Genetics*, **17** (7) 407-413

Erlich. H.A., Gelfand, D., Sninsky, J.J. (1991) Recent advances in the polymerase chain reaction. *Science*, **252** 1643-1650

Feder, J.N., Tsuchihashi, Z., Irrinki, A., Lee, V.K., Mapa, F.A., Morikang, E. (1997) The haemochromatosis founder mutation in HLA-H disrupts beta2-microglobulin interaction and cell surface expression. *Journal of Biological Chemistry*, **272** 14025-14028

Förster, T. (1946) Energuewanderung und Fluoreszenz. *Naturwissenschaften*, **6** 166-175

Förster, T. (1948) Zwishenmoleculare Energiewanderung und Fluoreszenz. *Annals of Physics (Leipzig)*, **2** 55-75

Frackman, S., Kobs, G., Simpson, D., Storts, D. (2001) Betaine and DMSO: Enhancing agents for PCR. *Promega Notes*, **65** 21-27

Freeland, S.J. Hurst, L.D. (1998) The genetic code is one in a million. *Journal of Molecular Evolution*, **47** (3) 238-248

French, D.J., Archard, C.L., Brown, T., McDowell, D.G. (2001) HyBeacon® probes: a new tool for DNA sequence detection and allele discrimination. *Molecular and Cellular Probes*, **15** 363-374

Fujii, K., Matsubara, Y., Akanuma, J., Takahashi, K., Kure, S., Suzuki, Y., Imaizumi, M., Iinuma, K., Sakatsume, O., Rinaldo, P., Narisawa, K. (2000) Mutation detection by TaqMan-Allele specific amplification: Application to molecular diagnosis of glycogen storage disease Type Ia and medium-chain acyl-CoA dehydrogenase deficiency. *Human Mutation*, **15** 189-196

Furuya, H., Fernandez-Salguero, P., Gregory, W. (1995) Genetic polymorphism of CYP2C9 and its effect on warfarin maintenance dose requirement in patients undergoing anticoagulation therapy. *Pharmacogenetics*, **5** 389-392

Geever, R.F., Wilson, L.B. Nallaseth, F.S., Milner, P.F., Bittner, M., Wilson, J.T. (1981) Direct amplification of sickle cell anaemia by blot hybridisation. *Proceedings of the National Academy of Science USA*, **78** 5081-5085

Germer, S., Higuchi, R. (1999) Single-tube genotyping without oligonucleotide probes. *Genome Research*, **9** 72-78

Germer, S., Holland, M.J., Higuchi, R. (2000) High-Throughput SNP Allele-Frequency Determination in Pooled DNA Samples by Kinetic PCR. *Genome Research*, **10** 258-266

Gibbs, R.A., Nguyen, P., and Caskey, C.T. (1989) Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Research*, **17** (7) 2437-2448

Giguére, Y. Rousseau, F. (2000) The genetics of osteoporosis: 'complexities and difficulties'. *Clinical Genetics*, **57** 161-169

Ginsburg, G.S., McCarthy, J.J. (2001) Personalised medicine: revolutionising drug discovery and patient care. *Trends in Biotechnology*, **19** (12) 491-496

Glaab, W.E., Skopek, T.R. (1999) A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan®) and mismatch amplification mutation assay. *Mutation Research*, **430** 1-12

Goodman, M.F., Creighton, S., Bloom, L.B., and Petruska, J. Biochemical Basis of DNA Replication Fidelity (1993) *Critical reviews in Biochemistry and Molecular Biology*, **28** (2) 83-126

Gray, I.C., Campbell, D.A., Spurr, N.K. (2000) Single nucleotide polymorphisms as tools in human genetics. *Human Molecular Genetics*, **9** (16) 2403-2408

Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y., Barringer, K.J., Richman, D.D., Gingeras, T.R. (1990) Isothermal, in vitro amplification of nucleic acid by a multienzyme reaction modelled after retroviral replication. *Proceedings of the National Academy of Science USA*, **87** 1874-1878

Hamajima, N. (2001) PCR-CTTP: a new genotyping technique in the era of genetic epidemiology. *Expert Review of Molecular Diagnostics*, **1** (1) 119-123

Henegariu, O., Heerema, N.A., Dlouhy S.R., Vance, G.H. and Vogt, P.H. (1997) Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *BioTechniques*, **23** 504-511

Hessner, M.J., Budish, M.A., Friedman, KD. (2000) Genotyping of Factor V G1691A (Leiden) without the use of PCR by invasive cleavage of oligonucleotide probes. *Molecular Diagnostics and Genetics*, **46** (8) 1051-1056

Hézard, N., Cornillet, P., Droullé, C., Gillot, L., Potron, G., Nguyen, P. (1997) Factor V leiden: Detection on whole blood by ASA PCR using an additional mismatch in antepenultimate position. *Thrombosis Research*, **88** (1) 59-66

Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, G (1992) Simultaneous amplification and detection of specific DNA sequences. *Bio-Technology*, **10** 413-417

Higuchi, R., Fockler, C., Dollínger, G., Watson, R. (1993) Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Bio-Technology*, **1 1** 1026-1030

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H. (1991) Detection of specific polymerase chain reaction product by utilising the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Science USA*, **88** 7276-7280

Holloway, A.J., Laar, R.K., Tothill, R.W., Bowtell, D.D.L. (2002) Option available-from start to finish- for obtaining data from DNA microarrays II. *Nature Genetics Supplement,* **32** 481-489

Howell, W.M., Jobs, M., Gyllensten, U., Brookes, A.J. (1999) Dynamic Allele Specific Hybridisation. A new method for scoring single nucleotide polymorphisms. *Nature Biotechnology*, **17** 87-88

Huang, M-M., Arnheim, N., Goodman, M.F. (1992) Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR, *Nucleic Acids. Research*, **20** 4567-4573

Humphries, S.E., Talmud, P.J., Hawe, E., Bolla, M., Day, I.N.M., Miller, G.J. (2001) Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study. *The Lancet*, **358** 115-119

Ingram, V.M. (1957) Gene mutation in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*, **180** 326-328

Jackson, P.E., Scholl, P.F., Groopman, J.D. (2000) Mass spectrometry for genotyping: an emerging tool for molecular medicine. *Molecular Medicine Today*, **6** 271-276

Jensen, G.A., Singh, S.K., Kumar, R., Wengel, J., Jacobsen, J.P. (2001) A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex. *Journal of the Chemical Society*, **2** 1224-1232

Kainz, P. (2000) The PCR plateau phase – towards an understanding of its limitations. *Biochimica et Biophysica Acta*, **1494** 23-27

Kainz, P., Schmiedlechner, A., Strack, B. (2000) Specificity enhanced Hot-Start PCR: Addition of double stranded DNA fragments adapted to the annealing temperature. *BioTechniques*, **28** 278-282

Kan, Y.W., Dozy, A.M., Alter, B.P., Frigoletto, F.D., Nathan, D.G. (1972) Detection of the sickle cell gene in the human fetus: potential for intrauterine diagnosis of sickle cell anaemia. *New England Journal of Medicine*, **287** 1-5

Kelley, S.O., Boon, E.M., Barton, J.K., Jackson, N.M., Hill, M.G. (1999) Single-base mismatch detection based on charge transduction through DNA. *Nucleic Acids Research*, **27** (24) 4830-4837

Kellogg, D E, Rybalkin, I, Chen, S, Mukhamedova, N, Vlasik, T, Siebert, P D, Chenchik, A (1994) TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against *Taq* DNA polymerase *BioTechniques*, **16** (6) 1134-1137

Kim, S.-J, Venstra-VanderWeele J., Hanna, G.L., Gonen, D., Leventhal, B.L., Cook, E.H. (2000) Mutation screening of human 5-HT(2B) receptor gene in early-onset obsessive compulsive disorder. *Molecular and Cellular Probes*, **14** 47-52

Klein, D. (2002) Quantification using real-time PCR technology: Application and limitations. *Trends in Molecular Medicine*, **8** (6) 257-260

Koskin, A.A., Singh, S.K., Nielsen, P., Rajwanshi, V.K., Kumar, R., Meldgaard, M. (1998) LNA (locked nucleic acids): synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, **54** 3607-3630

Kovárová, M., Dáraber, P. (2000) New specificity and yield enhancer of polymerase chain reactions. *Nucleic Acids Research*, **28** (13) e70

Kreader, C. (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology*, **62** (3) 1102-1106

Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, **22** 139-144

Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A., and Benkovic, S.J. (1987) Kinetic Mechanism of DNA Polymerase I (Klenow). *Biochemistry*, **26** 8410-8417

Kwok, P-Y. (2001) Methods for genotyping single nucleotide polymorphisms. *Annual Review Genomics and Human Genetics*, **2** 235-258

Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C., Sninsky, J. J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, **18** (4) 999-1005

Lander, E.S. Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409** 860-921

Lay, M.J., Wittwer, C.T. (1997) Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. *Clinical Chemistry*, **43** (12) 2262-2267

Lee, H.H. (1996) Ligase Chain Reaction. *Biologicals*, **24** 197-199

Lee, M.A., Siddle, A.L., Page, R.H. (2001) ResonSense®: simple linear fluorescent probes for quantitative homogenous rapid polymerase chain reaction. *Analytica Chimica Acta*, **21764** 1-10

Lewin, B. (1995) Genes V. Oxford University Press, Walton Street, Oxford, UK.

Liljedahl, U., Syvänen, A-C. (2002) SNP genotyping: current methods and practical applications. *Clinical Laboratory international*, May Edition, pp 6-9

Lin, Y., Jayasena, S.D. (1997) Inhibition of multiple thermostable DNA polymerases by a heterodimeric aptamer. *Journal of Molecular Biology*, **271** 100-111

Linz, U., Delling, U., Rubsamen-Waigmann, H. (1990) Systematic studies on Parameters Influencing the Performance of the Polymerase Chain Reaction. *Journal of Clinical Chemistry and Clinical Biochemistry*, **28** 5-13

Li, H., Cui, X., Arnheim, N. (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. *Proceedings of the National Academy of Science USA*, **87** 4580-4584

Lipshutz, R.J., Morris, D., Chee. M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R., Fodor, S.P.A. (1995) Using oligonucleotide probe arrays to access genetic diversity. *BioTechniques*, **19** (3) 442-447

Liu, W., Saint, D.A. (2002) Validation of a quantitative method for real time PCR kinetics. *Biochemical and Biophysical Research Communications*, **294** 347-353

Livak (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genetic Analysis*, **14** 143-149

Livak, K.J., Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*, **25** 402-408

Lizardi P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genetics*, **19** 225-232

Lu, Y-H., Negre, S. (1993) Use of glycerol for enhanced efficiency and specificity of PCR amplification. *Trends in Genetics*, **9** (9) 297

Lui, Q., Thorland, E.C., Heit, J.A., Sommer, S.S. (1997) Overlapping PCR for Bi-directional PCR amplification of specific alleles: A rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. *Genome Research*, **7** 389-398

Lundin, A., Rickardsson, A., Thore, A. (1976) Continuous monitoring of ATP-converting reactions by purified firefly luciferase. *Analytical Biochemistry*, **75** 611-620

Lunn, G. Sansone, E.B. (1987) Ethidium bromide: destruction and decontamination of solutions. *Analytical Biochemistry*, **162** 453-458

Lyamichev, V., Mast, A.L., Hall, J.G., Prudent, J.R., Kaiser, M.W., Takova, T. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology*, **17** 292-296

Markovits, J., Roques, B.P., Le Pecq, J.B. (1979) Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Analytical Biochemistry*, **94** 259-264

Mason, W.T. (1999) Fluorescent and Luminescent Probes for Biological Activity. A Practical Guide to Technology for Quantitative Real-Time Analysis, Second Ed., Academic Press, pp. 647

Matsakis, M., Berdoukas, V.A., Angastiniotis, M. (1980) Haemotological aspects of antenatal diagnosis for thalassaemia in Britain. *British Journal of Haematology*, **45** 185-197

Matsubara, Y., Fujii, K., Rinaldo, P., Narisawa, K., (1999) A fluorogenic allele-specific amplification method for DNA-based screening for inherited metabolic disorders. *Acta Paediatra (Suppl)*, **88** 65-68

McCarthy, J.J., Hilfiker, R. (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nature*, **18** 505-508

Mei, R., Glipeau, P.C., Prass, C., Berno, A., Ghandour, G. (2000) Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Research*, **10** 1126-1137

Mendelman, L.V., Petruska, J., Goodman, M.F. (1990) Base mispair extension kinetics: Comparison of DNA polymerase a and reverse transcriptase, *Journal of Biological Chemistry*, **265** 2338-2346

Michaelis, L., Menten, M. (1913) Die kinetik der invertinwirkung. *Biochemische Zeitschrift,* **49** 333-369

Mizugaki, M., Hiratsuka, M., Agatsuma, Y., Matsubara, Y., Fujii, K., Kure, S., Narisawa, K. (2000) Rapid detection of CYP2C18 genotypes by real-time fluorescence polymerase chain reaction. *Journal of Pharmacy and Pharmacology*, **52** 199-205

Modiano, D. (1996) Different response to *Plasmodium falciparum* malaria in west African sympatric ethnic groups. *Proceedings of the National Academy of Science USA*, **93** 13206-13211

Morgan, T.H. (1910) Sex-limited inheritance in Drosophila. *Science*, **32** 120-122

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold Spring Harbor Simple Quantitative Biology, **51** 263-273

Murayama, M. (1966) Tertiary structure of sickle cell hemoglobin and its functional significance. *Journal of Cell Physiology*, **67** (3) 21-32

Myakishev, M.V., Khripin, Y., Hu, S., Hamer, D.H. (2001) High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labelled primers. *Genome Research*, **11** 163-169

Nagai, M., Yoshida, A., Sato, N. (1998) Additive effects of bovine serum albumin, dithiothreitol, and glycerol on PCR. *Biochemistry and Molecular Biology International,* **44** (1) 157-163

Nath, K., Sarosy, J.W., Hahn, J., Di Como, C. (2000) Effects of ethidium bromide and SYBR® Green I on different polymerase chain reaction systems. *Journal of Biochemical and Biophysical Methods,* **42** 15-29

Neumaier, M., Braun, A., Wagener, C. (1998) Fundamentals of quality assessment of molecular amplification methods in clinical diagnostics. *Clinical Chemistry*, **44** (1) 12-26

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., and Markham, A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research*, **17** 2503-2516

Nichols, W.C., Liepnicks, J.J., McKusick, V.A., Benson, M.D. (1989) Direct sequencing of the gene for Maryland/German familial amyloidotic polyneuropathy type II and genotyping by allele-specific enzymatic amplification. *Genomics*, **5** 535-540

Nielsen, P.E., Egholm, M., Berg, R.H., Buchard, O. (1991) Sequence selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, **254** 1491-1500

Nikiforov, T.T., Rendle, R.B., Goelet, P., Rogers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L., Knapp, M.R. (1994) Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Research*, **22** 4167-4175

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., Hase, T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research,* **28** (12) e63

Nowotny, P., Kwon, J.M., Goate, A.M. (2001) SNP analysis to dissect human traits. *Current Opinion in Neurobiology*, **11** 637-641

Old, J.M., Petrou, M., Varnavides, L., Layton, M., Modell, B. (2000) Accuracy of prenatal diagnosis for haemoglobin disorders in the UK: 25 years' experience. *Prenatal Diagnosis*, **20** 986-991

Old, J.M., Varawalla, N.Y., Weatherall, D.J. (1990) The rapid detection and prenatal diagnosis of β thalasseamia in the Asian Indian and Cypriot populations in the UK. *The Lancet*, **336** 834-837

Orum, H., Jakobsen, M.H., Koch, T., Vuust, J., Borre, M.B. (1999) Detection of the Factor V Leiden Mutation by direct allele-specific hybridisation of PCR amplicons to photoimmobilised Locked Nucleic Acids. *Clinical Chemistry*, **45** (11) 1898-1905

Orum, H., Nielsen, P.E., Egholm, M., Berg, R.H., Buchart, C., Stanley, C. (1993) Single base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Research,* **24** 983-984

Pangas, S., Woodruff, T. (2002) Genetic Engineering News, March Edition, pp. 12

Peccoud, J. and Jacob, C. (1998) Statistical Estimations of PCR Amplification Rates. Gene Quantification, Birkhauser, Boston

Peirce, J. (2002) High throughput thermocyclers. *The Scientist*, **16** (22) Nov 11

Perrin, F. (1926) Polarisation de la lumiere de fluorescence. Vie moyenne de molecules dans l'etat excite. *Journal de Physique Radium*, **7** 390-401

Petruska, J., Goodman, M.F., Boosalis, M.S., Sowers, L.S., Cheong, C., Tinoco, I. (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity, *Proceedings of the National Academy of Science USA*, **85** 6252-6256.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, **29** 9

Pirmohamed, M., Park, K. (2001) Genetic susceptibility to adverse drug reactions. *Trends in Pharmacological Sciences*, **22** (6) 298-305

Raeymaekers, L. (1993) Quantitative PCR: Theoretical considerations with practical implications. *Analytical Biochemistry*, **214** 582-585

Rasmussen, R., Morrison, T., Herrman, M., Wittwer, C. (1998) Quantitative PCR by continuous monitoring of a double strand specific binding dye. *Biochemica*, **2** 8-12

Ririe, K.M., Rasmussen, R., Wittwer, C.T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry*, **245** 154-160

Ronaghi, M., Karamohamed, S., Petterson, B., Uhlen, M., Nyren, P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, **242** 84-89

Ross, P., Hall, L., Smirnov, I., Haff, L. (1998) High-level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnology*, **16** 1347-1351

Rychlik, W. (1995) Priming efficiency in PCR. *BioTechniques*, **18** (1) 84-89

Rychlik, W., Spencer, W.J., Rhoads, R.E. (1990) Optimisation of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Research*, **18** (21) 6409-6412

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230** 1350-1354

Saiki, R.K., Walsh, P.S. Levenson, C.H., Erlich, H.A. (1989) Genetic analysis of amplified DNA with immobilised sequence-specific oligonucleotide probes, *Proceedings of the National Academy of Science USA*, **86** 6230-6234

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press

Sanders, R. (2002) Testing times. *Expert Review of Molecular Diagnostics*, **2** (6) 519-521

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science USA*, **74** 5463-5467

Sarkar, G., Cassady, J., Bottema, C.D.K., Sommer, S.S. (1990) Characterisation of polymerase chain reaction amplification of specific alleles. *Analytical Biochemistry,* **186** 64-68

Sasvari-Szekely, M., Gerstner, A., Ronai, A., Staub, M., Guttman, A. (2000) Rapid genotyping of factor V Leiden mutation using single tube bi-directional allele-specific amplification and automated ultrathin-layer agarose gel electrophoresis. *Electrophoresis*, **21** 816-821

Sauer, S., Lachner, D, Berlin, K., Lehrach, H., Escary, J.L., Fox, N., Gut, I. G. (2000) A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Research*, **28** e13

Schildkraut, C., Lifson, S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, **3** 195-208

Schnell, S., Mendoza, C. (1997A) Closed form solution for time-dependent enzyme kinetics. *Journal of Theoretical Biology*, **187** 207-212

Schnell, S., Mendoza, C. (1997C) Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction (QC-PCR), *Journal of Theoretical Biology*, **184** 433-440

Schnell, S., Mendoza, C. (1997B) Theoretical description of the Polymerase Chain Reaction, *Journal of Theoretical Biology*, **188** 313-318

Schumaker, J.M., Metspalu, A., Caskey, C.T. (1996) Mutation detection by solid-phase primer extension. *Human Mutation*, **7** 346-354

Seggern, D.H.V. (1993) CRC Standard Curves and Surfaces, CRC Press.

Serjeant, G.R. (1997) Sickle-cell disease. *The Lancet*, **350** 725-730

Sevall, J.S. (2000) Factor V Leiden genotyping using real-time fluorescent polymerase chain reaction. *Molecular and Cellular Probes*, **14** 249-253

Sharp, P.A., Sudgen, B., Sambrook, J. (1973) Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry*, **12** 3055-3063

Shen, W-H, Hohn, B., Miescher, F. (1992) DMSO improves PCR amplification of DNA with complex secondary structure. *Trends in Genetics*, **8** (7) 227

Singer, V.L., Jin, X., Jones, L., Yue, S., Haugland, R. (1997) Sensitive fluorescent stains for detecting nucleic acids in gels and solutions. United Medical Press.

Singh, S.A., Nielsen, P. Koshkin, A.A., Wengel, J. (1998) LNA (locked nucleic acid): synthesis and high affinity nucleic acid recognition. *Chemical Communications*, 455-456

Sommer, S.S., Groszbach, A.R., Bottema, C.D.K. (1992) PCR Amplification of Specific Alleles (PASA) is a general method for rapidly detecting known single-base changes. *BioTechniques*, **12** (1) 1152

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, **98** 503-517

Stolovitzky, G., Cecchi, G. (1996) Efficiency of DNA replication in the polymerase chain reaction, *Proceedings of the National Academy of Science USA*, **93** 12947-12952

Stryer, L., Haughland, R.P. (1967) Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Science USA*, **58** (2) 719-726

Syvänen, A-C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, **2** 930-939

Taguchi, G. (1986) Introduction to quality engineering. Asian Productivity Organisation, UNIPUB, New York

Takara Shuzo Co. (2000) www.takara.co.jp

Täpp, I., Malmberg, L., Rennel, E., Wik, M., Syvänen, A.-C. (2000) Homogenous scoring of single nucleotide polymorphisms: Comparison of the 5' nuclease TaqMan® assay and molecular beacon probes. *BioTechniques*, **28** 732-738

Taylor, J.G., Choi, E-H, Foster, C.B., Chanock, S.J. (2001) Using genetic variation to study human disease. *Trends in Molecular Medicine*, **7** (11) 2001

The International SNP MAP Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409** 928-933

The International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409** 860-921

Thelwell, N., Millington, S., Solinas, A., Booth, J., and Brown, T. (2000) Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Research*, **28** 3752-3761

Tonisson, N., Kurg, A., Kaasik, K., Lohmussaar, E., Metspalu, A. (2000) Unravelling genetic data by arrayed primer extension. *Clinical Chemistry and Laboratory Medicine*, **38** (2) 165-170

Townsend, A., Drakesmith, H. (2002) Role of *HFE* in iron metabolism, hereditary haemochromatosis, anaemia of chronic disease, and secondary iron overload. *The Lancet*, **359** 786-789

Tuma, R.S., Beaudet, M.P., Jin, X., Jones, L.J., Cheung, C-Y., Yue, S., Singer, V.L. (1999) Characterisation of SYBR Gold Nucleic Acid Gel Stain: A dye optimised for use with 300-nm ultraviolet transilluminators. *Analytical Biochemistry*, **268** 278-288

Tyagi, S., Bratu, D.P., Kramer, F.R. (1998) Multicolor molecular beacons for allele discrimination. *Nature Biotechnology*, **16** 49-53

Tyagi, S., Marras, S.A.E., Kramer, F.R. (2000) Wavelength-shifting molecular beacons. *Nature Biotechnology*, **18** 1191-1196

Ugozzoli, L.A., Chinn, D., Hamby, K. (2002) Fluorescent multicolour multiplex homogenous assay for the simultaneous analysis of the two most common haemochromatosis mutations. *Analytical Biochemistry*, **307** 47-53

Velikanov, M.V., Kapral, R. (1999) Polymerase Chain Reaction: A Markov process approach. *Journal of Theoretical Biology*, **201** 239-249

Venter, C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001) The sequence of the human genome. *Science*, **291** 1304-1351

Voorberg, J., Roesle, J., Koopman, R., Buller, H., berends, F., TenCate, J.W. (1994) Association of idiopathic venous thromboembolism with single point mutation at Arg506 of factor V. *The Lancet*, **343** 1535-1536

Walburger, D.K., Afonina, I.A., Wydro, R. (2001) An improved real time PCR method for simultaneous detection of C282Y and H63D mutations in the HFE gene associated with hereditary haemochromatosis. *Mutation Research Genomics*, **432** 69-78

Watson, J.D., Crick, F.H.C. (1953) Molecular structure of nucleic acids, *Nature*, **171** 737-738, *and* Genetical implications of the structure of deoxyribonucleic acid, *Nature*, **171** 964-967

Weiss, G., Haeseler, A. (1997) A coalescent approach to the polymerase chain reaction, *Nucleic Acids Research*, **25** 3082-3087

Wharam, S.D., marsh, P., Lloyd, J.S., Ray, T.D., Mock, G.A., Assenberg, R., McPhee, J., Brown, P., Weston, A., Cardy, D.L.N. (2001) Specific detection of DNA and RNA targets using a novel isothermal nucleic acid amplification assay based on the formation of a three-way junction structure. *Nucleic Acids Research*, **29** (11) e54

Wharton, C.W., Eisenthal, R. (1981) Molecular Enzymology, Blackie and Son Ltd.

Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., Little, S. (1999) Detection of PCR products using self-probing amplicons and fluorescence. *Nature Biotechnology*, **17** 804-807

Whiteley, N.M., Hunkapillar, M.W., Glazer, A.N. (1989) Detection of specific sequences in nucleic acids. In US Patent No. 4883750 (Applied Biosystems)

Wieczorek, S.J. Tsongalis, G.J. (2001) Pharmacogenomics: will it change the field of medicine? *Clinica Chimica Acta*, **308** 1-8

Wilson, P.W., Schaefer, E.J., Larson, M.G., Ordovas, J.M. (1996) Apolipoprotein E alleles and risk of coronary disease: a meta-analysis. *Arteriosclerosis Thrombosis and Vascular Biology*, **16** 1250-1255

Wittwer, C.T., Garling, D.J. (1991) Rapid cycle DNA amplification: Time and temperature optimization. *BioTechniques*, **10** 76-83

Wittwer, C.T., Herrmann, M.G., Moss, A.A., Rasmussen, R.P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification, *BioTechniques*, **22** 130-138

Wittwer, C.T., Marshall, B.C., Reed, G.H., and Cherry, J.L. (1993) Rapid Cycle Allele-Specific Amplification: Studies with the Cystic Fibrosis ΔF508 Locus. *Clinical Chemistry*, **39** 804-809

Wolffs, P., Knutsson, R., Sjoback, R., Radstrom, P. (2001) PNA-based light-up probes for real-time detection of sequence-specific PCR products. *BioTechniques*, **31** 766-771

Wong, I., Patel, S.S., Johnson, K.A. (1991) An induced-fit kinetic mechanism for DNA replication fidelity: Direct measurement by single-turnover kinetics, *Biochemistry*, **30** 526-537

Wu, D.Y., Ugozzoli, L., Pal, B.K., Qian, J., Wallace, R.B. (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA and Cell Biology*, **10** (3) 233-238

Wu, D.Y., Ugozzoli, L., Pal, B.K., Wallace, B. (1989) Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National Academy of Science USA*, **86** 2757-2760

Ye, S., Dhillon, S., Ke, X., Collins, A.R., and Day, I.N.M. (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*, **29** e88

Ye, S., Humphries, S., and Green, F. (1992) Allele specific amplification by tetra-primer PCR. *Nucleic Acids Research*, **20** 1152

Appendices

Appendix I

Primer Sequences for PCR Amplifications

Oligo Name	*#bp*	T_m[1]	*Sequence (5'-3')*
ACT BF	23	63.2	TTC CGT AGG ACT CTC TTC TCT GA
ACT BR	23	67.0	GGG GTG TTG AAG GTC TCA AAC AT
ACT BR A	34	81.0	CGG GAT CCA GAG GGG TGT TGA AGG TCT CAA ACA A
ACT BR C	34	83.1	CGG GAT CCA GAG GGG TGT TGA AGG TCT CAA ACA C
ACT BR G	34	83.1	CGG GAT CCA GAG GGG TGT TGA AGG TCT CAA ACA G
ACT BR T	34	81.0	CGG GAT CCA GAG GGG TGT TGA AGG TCT CAA ACA T
ASA 1A	15	59.5	TCG ATT CGG GAT CCA
ASA 1C	15	60.5	TCG ATT CGG GAT CCC
ASA 1G	15	61.9	TCG ATT CGG GAT CCG
ASA 1T	15	57.1	TCG ATT CGG GAT CCT
C282Y MUT AS	18	66.5	CTG GGT GCT CCA CCT GGT
C282Y MUT OP	20	67.3	GGG CTC CCA GAT CAC AAT GA
C282Y MUT OP1	22	64.8	GCC AAG GAG TTC GAA CCT AAA G
C282Y WT AS	23	64.7	GGG GAA GAG CAG AGA TAT ACG TG
C282Y WT OP	22	64.8	GGG CTC CCA GAT CAC AAT GA
H63D MUT AS	19	67.1	CCA CAC GGC GAC TCT CAT C
H63D MUT OP	21	65.5	GCA GGA CCT TGG TCT TTC CTT
H63D WT AS	23	64.6	CCA GCT GTT CGT GTT CTA TGA TC
H63D WT OP	22	66.0	ATC CCA CCC TTT CAG ACT CTG A
ICF	21	65.1	TGG CAG CAC TGC ATA ATT CTC
ICR	21	60.8	AGC GGT AAG ATC CTT GAG AGT
Oligo Name	bp	T_m	Sequence (5'-3')
MUT OP 267	21	61.8	GGG TTT GAA GTC CAA CTC CTA
PM AS A	19	64.7	GAA TTC GAT TCG GGA TCC A
PM AS C	19	65.3	GAA TTC GAT TCG GGA TCC C
PM AS G	19	66.4	GAA TTC GAT TCG GGA TCC G
PM AS T	19	62.7	GAA TTC GAT TCG GGA TCC T
SC MUT AS	22	67.5	GCA GTA ACG GCA GAC TTC TCC A
SC MUT OP	20	65.3	AGG GCA GAG CCA TCT ATT GC
SC WT AS	20	65.9	ATG GTG CAC CTG ACT CCT GA
SC WT OP	20	67.5	AGG GCC TCA CCA CCA ACT TC
WT CP 98	17	64.1	GCT GTC CCC AGT GGC TT
WT CP 374	31	76.3	CGG GAT CCT TCC GTA GGA CTC TCT TCT CTG A
WT OP 517	21	61.7	CCC CTT CCT ATG ACA TGA ACT
WT FAM	27	62.7	FAM AGA AGT CTG CCG TTA CTG CCC TGT GGT TAMRA
MUT JOE	27	62.7	JOE AGG AGT CAG GTG CAC CAT GGT GTC TGT TAMRA

[1] T_m values are estimated at 0.1M Na^+

Appendix II

pGEM-T® Easy Vector System II plasmid sequence with cloned insert

Cloned b-actin gene insert displayed in *italic type* (position 52), mutation site displayed in ***bold italic type*** (position 60), position of internal control fragment shown in **bold type** (position 1820 to 2047)

1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCATGGCGGCCGCG

51 GGAATTCGATT- (cloned insert)

*CGGGATCC**G**GAGGGGTGTTGAAGGTCTCAAACATGATCTGTAAGGCAGAGATGCACCATGT*

CACACTGGGGAAGCCACTGGGGACAGCCAGGCCAGACGGGGGACATGCAGAAAGTGCAAA

GAACACGGCTAAGTGTGCTGGGGTCTTGGGATGGGGAGTCTGTTCAGACCTACTGTGCGCC

TACTTAATACACACTCCAAGGCCGCTTTACACCAGCCTCATGGCCTTGTCACACGAGCCAGT

GTTAGTACCTACACCCACAACACTGTCTCAGACACCTAGTCAGAGAGACAAACACCAGAAAA

AGAGCTCATCTGGGAAAAAGCAAATAGAACCTGCAGAGTTCCAAAGGAGACTCAGGTCAGA

GAAGAGAGTCCTACGGAAGGATCCCGAA- (end of cloned insert)

52 ATCACTAGTG AATTCGCGGC CGCCTGCAGG TCGACCATAT

101 GGGAGAGCTC CCAACGCGTT GGATGCATAG CTTGAGTATT CTATAGTGTC

151 ACCTAAATAG CTTGGCGTAA TCATGGTCAT AGCTGTTTCC TGTGTGAAAT

201 TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG

251 TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC

301 GCTCACTGCC CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA

351 TGAATCGGCC AACGCGCGGG GAGAGGCGGT TTGCGTATTG GGCGCTCTTC

401 CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG

451 CGGTATCAGC TCACTCAAAG GCGGTAATAC GGTTATCCAC AGAATCAGGG

501 GATAACGCAG GAAAGAACAT GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA

551 CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC CGCCCCCCTG

601 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA

651 GGACTATAAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC

701 TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT

751 CGGGAAGCGT GGCGCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG

801 GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA

851 GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG

901 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC

951 AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA

1001 CTACGGCTAC ACTAGAAGAA CAGTATTTGG TATCTGCGCT CTGCTGAAGC

1051 CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC

1101 ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA TTACGCGCAG

1151 AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG

1201 CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GAGATTATCA

1251 AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC

1301 AATCTAAAGT ATATATGAGT AAACTTGGTC TGACAGTTAC CAATGCTTAA

1351 TCAGTGAGGC ACCTATCTCA GCGATCTGTC TATTTCGTTC ATCCATAGTT

1401 GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG GCTTACCATC

1451 TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CCGGCTCCAG

1501 ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT

1551 CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC

1601 TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT GTTGCCATTG

1651 CTACAGGCAT CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTCAGC

1701 TCCGGTTCCC AACGATCAAG GCGAGTTACA TGATCCCCCA TGTTGTGCAA

1751 AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG

1801 CCGCAGTGTT ATCACTCATG **GTTATGGCAG CACTGCATAA TTCTCTTACT**

1851 GTCATGCCAT CCGTAAGATG CTTTTCTGTG ACTGGTGAGT ACTCAACCAA

1901 GTCATTCTGA GAATAGTGTA TGCGGCGACC GAGTTGCTCT TGCCCGGCGT

1951 CAATACGGGA TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC

2001 ATTGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT TACCGCTGTT

2051 GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA TCTTCAGCAT

2101 CTTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT

2151 GCCGCAAAAA AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT

2201 CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA

2251 GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTCCG

2301 CGCACATTTC CCCGAAAAGT GCCACCTGAT GCGGTGTGAA ATACCGCACA

2351 GATGCGTAAG GAGAAAATAC CGCATCAGGA AATTGTAAGC GTTAATATTT

2401 TGTTAAAATT CGCGTTAAAT TTTTGTTAAA TCAGCTCATT TTTTAACCAA

2451 TAGGCCGAAA TCGGCAAAAT CCCTTATAAA TCAAAAGAAT AGACCGAGAT

2501 AGGGTTGAGT GTTGTTCCAG TTTGGAACAA GAGTCCACTA TTAAAGAACG

2551 TGGACTCCAA CGTCAAAGGG CGAAAAACCG TCTATCAGGG CGATGGCCCA

2601 CTACGTGAAC CATCACCCTA ATCAAGTTTT TTGGGGTCGA GGTGCCGTAA

2651 AGCACTAAAT CGGAACCCTA AAGGGAGCCC CCGATTTAGA GCTTGACGGG

2701 GAAAGCCGGC GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG

2751 GGCGCTAGGG CGCTGGCAAG TGTAGCGGTC ACGCTGCGCG TAACCACCAC

2801 ACCCGCCGCG CTTAATGCGC CGCTACAGGG CGCGTCCATT CGCCATTCAG

2851 GCTGCGCAAC TGTTGGGAAG GGCGATCGGT GCGGGCCTCT TCGCTATTAC

2901 GCCAGCTGGC GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG

2951 CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACGGCCA GTGAATTGTA

3001 ATACGACTCA CTATA

Appendix III

Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis

Nucleic Acids Research, 2001, Vol. 29, No. 23 e119

Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis

Christy M. Waterfall* and Benjamin D. Cobb

Molecular Sensing plc, Unit 3, Challeymead Business Park, Bradford Road, Melksham, Wiltshire SN12 8LH, UK

Received August 31, 2001; Revised and Accepted October 9, 2001

ABSTRACT

Allele-specific amplification (ASA) is a generally applicable technique for the detection of known single nucleotide polymorphisms (SNPs), deletions, insertions and other sequence variations. Conventionally, two reactions are required to determine the zygosity of DNA in a two-allele system, along with significant upstream optimisation to define the specific test conditions. Here, we combine single tube bi-directional ASA with a 'matrix-based' optimisation strategy, speeding up the whole process in a reduced reaction set. We use sickle cell anaemia as our model SNP system, a genetic disease that is currently screened using ASA methods. Discriminatory conditions were rapidly optimised enabling the unambiguous identification of DNA from homozygous sickle cell patients (HbS/S), heterozygous carriers (HbA/S) or normal DNA in a single tube. Simple downstream mathematical analyses based on product yield across the optimisation set allow an insight into the important aspects of priming competition and component interactions in this competitive PCR. This strategy can be applied to any polymorphism, defining specific conditions using a multifactorial approach. The inherent simplicity and low cost of this PCR-based method validates bi-directional ASA as an effective tool in future clinical screening and pharmacogenomic research where more expensive fluorescence-based approaches may not be desirable.

INTRODUCTION

The successful implementation of single nucleotide polymorphism (SNP) screening for clinical research and the administration of medicine tailored to an individual's genotype, so-called 'personalised medicine', lies in the ability to deliver highly reliable, accurate and inexpensive assays that can discriminate between key DNA polymorphisms. Biotechnology and pharmaceutical companies have addressed this challenge with the development of methodologies including single-strand conformation polymorphism analysis, heteroduplex analysis and denaturing gel electrophoresis. Each of these techniques relies on conformation-induced gel mobility changes (1). Simple homogeneous assays based on fluorescence resonance energy transfer, mass spectrometry or direct sequencing of PCR products have also been shown to be effective tools for SNP detection (2). Implementation of these methods can require specific reagent synthesis and downstream purification, making them complicated and expensive.

However, assays based on PCR for SNP diagnosis have broad potential in clinical diagnostics because of their inherent simplicity and potential low cost. Established PCR-based methodologies include certain ligation assays (3,4), genetic bit analysis (5), arrayed primer extension (6) and restriction fragment length polymorphisms (7), and use of sequence-specific fluorescent probes (8).

PCR has also been adapted for the detection of well-characterised SNPs using allele-specific amplification (ASA); oligonucleotides complementary to a given DNA sequence except for a mismatch at their 3′-hydroxyl residue will not function as primers in the PCR under appropriate conditions. A typical ASA test consists of two complementary reactions, each containing a common primer, an allele-specific primer and *Taq* DNA polymerase lacking 3′→5′ proofreading activity. The first reaction contains a primer specific for the normal (or wild-type) DNA sequence and refractory to amplification from mutant DNA at a given locus. Similarly, the second reaction contains a mutant-specific primer unable to amplify wild-type DNA. Molecular conformation is achieved by analysis of the resulting PCR amplicon profiles. A normal individual will generate product in the first reaction only; a heterozygote amplifies products in both reactions; and a homozygous mutant individual does so only in the mutant-specific reaction. Internal control amplification is necessary, providing a positive control for the PCR test (9).

More recently, single tube adaptations have been introduced for known SNP detection in a two-allele system including competitive oligonucleotide priming (10), multi-coloured fluorescent oligo labelling systems (2), tetra-primer PCR (11) and overlapping PCR strategies (12,13). The latter both use bi-directional primer arrangements in which common 'outer' primers define the size of each allele-specific fragment allowing simple identification using electrophoretic methods.

Maintaining the integrity of all ASA techniques, particularly for diagnostic application, requires the optimisation of many experimental parameters. The literature suggests that a magnesium and oligonucleotide titration is sufficient to obtain discriminatory conditions. Generally, further mismatches are incorporated at the 3′-end of the allele-specific primer to weaken hydrogen bonding between the primer and template, increasing the likelihood of discrimination (14). However, these conventional

*To whom correspondence should be addressed. Tel: +44 1225 706245; Fax: +44 1225 700056; Email: christy.waterfall@molecular-sensing.com

approaches are simplistic considering the complex component interactions in a PCR reaction and prove time consuming and expensive. This has limited the applicability of PCR-based SNP diagnosis for routine clinical application. In particular, bi-directional systems require significant optimisation due the effects of primer competition caused by multiple primer sets in the PCR amplification.

Here, we modify current systems at a number of levels with the application of a multifactorial optimisation procedure to single tube bi-directional ASA. Our strategy enables the rapid identification of component ratios favouring ASA in a greatly reduced experiment set using unmodified oligonucleotides. This low cost optimisation strategy combined with a single tube genotyping system introduces improvements to PCR-based SNP screening.

SNP model

We demonstrate our system by detecting the single base pair mutation that causes the autosomal recessive disease sickle cell anaemia (GenBank accession no. M34058). A single A→T transversion in the sequence encoding codon 6 of the human β-globin gene causes the substitution of amino acid glutamine by valine, forming a mutant globin chain termed HbS. Haemoglobin S is freely soluble when fully oxygenated, but under conditions of low oxygen tension the erythrocytes assume an irregular 'sickle' shape, leading to aggregation and haemolysis. The sickled erythrocytes become trapped in the microcirculation, depriving organs of essential oxygen, causing pain and chronic anaemia (15). Homozygous HbS is a serious haemoglobinopathy found almost exclusively in the Black population. About 8% of American Blacks are carriers and ~0.2% are affected. In the UK sickle cell disease affects ~10 000 people (actionresearch.co.uk/camp/sick.html). Heterozygotes (sickle cell carriers) are clinically normal, although their red cells will sickle when subjected to very low oxygen pressure *in vitro* (genelink.com/technical/sickle.html). PCR-based techniques for prenatal diagnosis of sickle cell disease were introduced in 1989 following the invention of the amplification refractory mutation system (9) and are still the method of choice in clinical research laboratories (16). A single tube adaptation of conventional ASA would add significant benefits to routine testing for sickle cell disease.

MATERIALS AND METHODS

Reagents

PCR reagents (*Taq* DNA polymerase, HS TaQUANT-OFF, 10× Mg-free *Taq* buffer, 25 mM $MgCl_2$) were obtained from Q-Biogene (CA), except for dNTPs (2 mM each dATP, dCTP, dGTP and dTTP), which were from Promega (UK). Primer selection was accomplished using the PrimerCalc© software package (Q-Biogene). For standard gel electrophoresis, molecular biology grade agarose was used (Q-Biogene) in 0.5× TBE buffer (45 mM Tris, 45 mM borate, 1 mM Na_2EDTA, pH 8.3). Double deionised water (18 MΩ) was used for all buffers and solutions for electrophoresis and PCR amplifications. Homozygous HbS human genomic DNA and heterozygous HbA/S DNA were a generous gift from the National Haemoglobinopathies Reference Service (Oxford, UK).

Bi-directional allele-specific PCR amplifications

WT-AS (5′-ATG GTG CAC CTG ACT CCT GA-3′) and WT-CP517 (5′-CCC CTT CCT ATG ACA TGA ACT-3′) were designed for amplification of a 517 bp fragment from the normal β-globin gene (wild-type primer set). MUT-AS (5′-CAG TAA CGG CAG ACT TCT CCA-3′) and MUT-CP267 (5′-GGG TTT GAA GTC CAA CTC CTA-3′) were designed for amplification of a 267 bp fragment from homozygous mutant DNA (HbS/S) conferring sickle cell disease (mutant primer set). All reactions were performed in a volume of 25 µl containing 10 ng total human genomic DNA template, 10× PCR reaction buffer, 1 U $µl^{-1}$ *Taq* DNA polymerase and water. The wild-type primer set, mutant primer set, $MgCl_2$ and dNTP concentrations are shown in Table 1 and were aliquoted according to an orthogonal array (Table 2) (17). Amplification was performed using an MJ Thermal Cycler (GRI, UK) through an initial denaturation at 95°C for 2 min, followed by 30 cycles of denaturation at 95°C for 30 s, primer annealing at 60°C for 30 s and extension at 72°C for 35 s. A small blind trial was performed in which 14 human genomic DNA samples of mixed zygosity were used as substrate DNA under discriminatory conditions that were considered optimal (reaction 9, Tables 1 and 2). The exact template concentration of each sample was not specified, varying between 5 and 25 ng $µl^{-1}$.

Analysis of PCR products

An aliquot (10 µl) of each PCR product was loaded with 6× loading buffer (Promega) onto a 2% agarose gel and run in 0.5× TBE buffer at 30 mA constant current. Amplification products were visualised by UV transillumination (305 nM)

Table 1. Concentration levels of components used in a multifactorial array for the identification of discriminatory reaction conditions for bi-directional ASA

Level	A	B	C
WT-AS/WT-CP517 (µM)	0.2	0.4	0.8
MUT-AS/MUT-CP267 (µM)	0.2	0.4	0.8
$MgCl_2$ (mM)	0.5	1	2
dNTPs (mM)	80	120	200

Table 2. An orthogonal array for testing four reaction parameters, each at three different levels

Reaction	WT-AS/ WT-CP517	MUT-AS/ MUT-CP267	$MgCl_2$	dNTPs
1	A	A	A	A
2	A	B	B	B
3	A	C	C	C
4	B	A	B	C
5	B	B	C	A
6	B	C	A	B
7	C	A	C	B
8	C	B	A	C
9	C	C	B	A

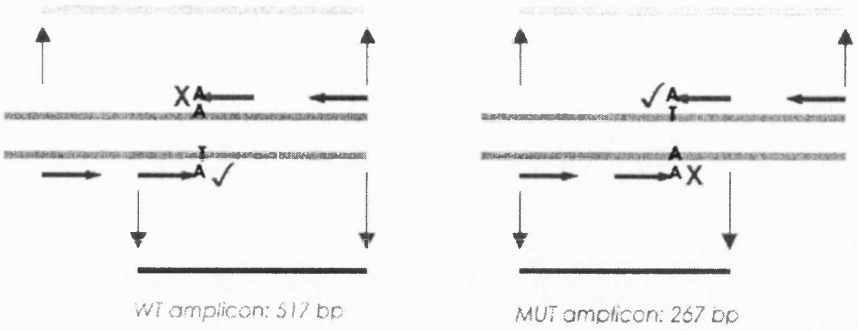


Figure 1. Schematic representation of bi-directional allele-specific PCR. The wild-type primer set is shown in green and the mutant primer set in pink. Three possible fragments can be formed depending upon substrate DNA zygosity, as well as enzyme system and cycling parameters. The green amplicon defines the allele-specific product formed when only the wild-type primers are extended and the pink amplicon defines the allele-specific product formed when only the mutant primers are extended. A heterozygote is identified by the presence of both PCR products. The target region amplified by non-allele-specific outer primers is shown in pale grey. Our selected enzyme system was inefficient at amplifying longer sequences and in the presence of competition from a smaller fragment insufficient target region was amplified for visualisation on a stained gel.

after staining with SYBR Gold (Molecular Probes, UK) for 30 min. The agarose gels were photographed and analysed by gel imaging software to obtain maximum band optical density measurements as a representation or 'score' of specific product yield.

RESULTS

Experimental design

A schematic of bi-directional ASA is shown in Figure 1. A wild-type allele-specific primer is designed as complementary to the *non-coding* strand of the wild-type allele terminating at the known polymorphic site. This primer will be extended in the presence of wild-type DNA only under appropriate PCR conditions. A further allele-specific primer is included in the reaction, designed as complementary to the *coding* strand of the known mutant allele, again terminating at the SNP site. This primer will be extended in the presence of the mutant allele only, as it is refractory to extension of wild-type DNA due to a single base pair mismatch at the 3′-end. Two 'outer primers' are placed on the opposite strand at a pre-defined number of bases from their allele-specific pair in order that molecular confirmation can be achieved by analysis of the resulting amplicon profiles using gel-based methods. A difference of at least 100 bp is preferable to increase electrophoretic separation.

Taq DNA polymerase will extend both outer primers in all reactions as they are non-allele-specific, but the yield of the large amplicon spanning the whole target region is subject to variation according to the enzyme system and cycling parameters employed. The HotStart enzyme system used in this study was less efficient at amplifying larger fragments and in the presence of competition from a smaller fragment did not amplify sufficient DNA for the larger band spanning the whole target region to be visualised on an agarose gel.

We designed two 20/21mer allele-specific primers, WT-AS and MUT-AS, complementary to the normal β-globin or sickle cell genes, respectively, arranged in a bi-directional orientation such that both primers terminate at the mutation site. Primer WT-AS has a single A nucleotide at the 3′-end, matching the T nucleotide on the non-coding strand of the normal β-globin gene. Primer MUT-AS has a single A nucleotide at the 3′-end, matching the T base on the coding strand of the sickle cell gene. The wild-type outer primer (WT-CP517) was positioned 517 bp downstream on the opposite strand of WT-AS. The mutant outer primer (MUT-CP267) was placed 267 bp upstream of MUT-AS. Unlike normal ASA, this assay has an inherent PCR control, as at least one allele-specific fragment should be yielded per reaction under optimised conditions. This negates the need for a distant pair of primers to direct the amplification of an internal control fragment.

Optimisation strategy for discrimination between the normal and sickle cell alleles

Inclusion of multiple allele-specific primer sets in a single reaction tube requires the consideration of many experimental factors. For typical multiplex reactions, the key parameter for optimisation is primer concentration. At equimolar concentrations, amplification by the two primer sets may be weak or strong at a given locus, an event dependent upon primer stabilities, binding efficiencies, priming competition, product size and concentration of other reaction components (18). Normally, primer ratios are individually optimised for each reaction, however, an adjustment in primer concentration has associated implications considering the complex interactions between other components in a PCR. Unique to this assay, each primer is complementary to a different part of the DNA sequence, introducing additional variability to the reaction.

$MgCl_2$ and dNTPs have been shown to affect the efficiency of priming and extension by altering the kinetics of association and dissociation of primer–template duplexes at annealing and extension temperatures. These components also alter the efficiency with which the polymerase recognises and extends such duplexes. Of particular relevance to allele-specific PCR methods, the concentration of $MgCl_2$ and dNTPs required for optimal amplification depends largely on the target and primer sequences. The presence of excess magnesium in a reaction may result in the accumulation of non-specific amplification products and insufficient concentrations reduce product yield.

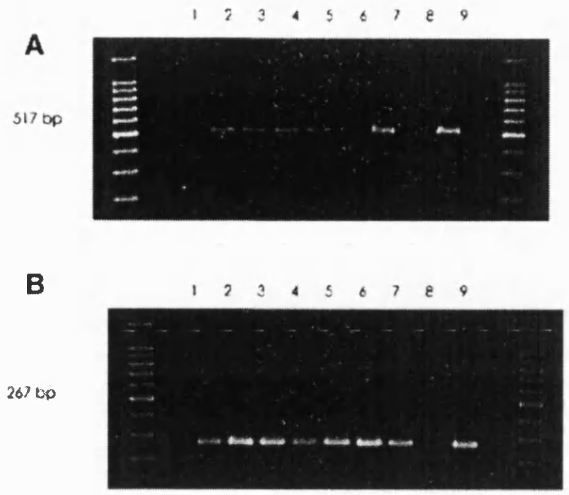


Figure 2. Gel image data of PCR amplifications. Outer lanes contain 100 bp ladders. Reactions 1–9 according to the matrix (Table 2) are labelled accordingly. Set A represents reactions in which wild-type template DNA was the substrate. Set B represents reactions in which HbS/S template DNA was included. The 517 and 267 bp allele-specific fragments were generated in each set, respectively. Product yield varied across each reaction set. Table 3 represents maximal optical density (OD) of the amplicon bands, used as a representation of product yield/reaction score.

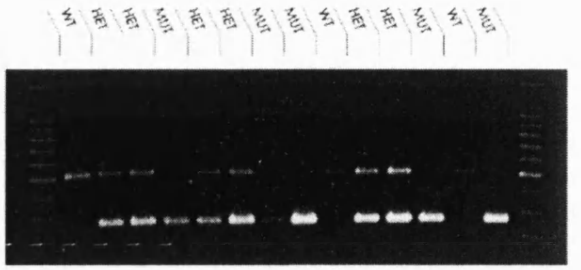


Figure 3. Gel image data from a small blind trial. Fourteen human DNA samples of varying zygosity were selected and amplified under conditions specified by reaction 9 (Table 2). Diagnosed template zygosity is shown on the gel image. Three normal samples were identified by amplification of the 517 bp product only. Six heterozygous samples were identified by amplification of both the 267 and 517 bp fragments. For heterozygotes, the 267 bp band was slightly more intense than the 517 bp band, likely due to staining disparity between optimisation sets A and B. Five homozygous sickle cell samples were identified by amplification of the 267 bp amplicon only.

It has also been demonstrated that dNTPs quantitatively bind Mg^{2+}, so that any modification of dNTP concentration will require a compensatory adjustment of $MgCl_2$ (17). Additionally, *Taq* polymerase requires free Mg^{2+} as a cofactor so any excess of dNTPS can have a detrimental effect on product yield by chelating cofactor ions (19). Standard approaches to multiplex or allele-specific PCR optimisation are unable to account for these multiple interactions, as they are largely based on trial-and-error strategies.

We applied a multifactorial optimisation strategy in which optimal component concentrations for allele discrimination were identified in a single trial with few reactions. The wild-type primer set, mutant primer set, $MgCl_2$ and dNTPs were selected as four critical parameters, the interactions of which are most likely to affect the outcome of bi-directional PCR. The concentrations of these components were varied according to an orthogonal array and differences in product yield across the reaction set were used for two purposes, qualitatively in the identification of diagnostic reaction conditions and quantitatively to study the individual effect of each parameter on the assay.

Optimal conditions for zygosity determination

Figure 2 shows gel image data from reaction sets A and B. Both sets are identical in component ratios and cycling conditions, containing normal and HbS/S DNA templates, respectively. Although product yield varied considerably across the reaction sets, only the appropriate allele-specific amplicon was generated according to template zygosity. Set A amplified the 517 bp amplicon and set B generated the 267 bp fragment only. The larger fragment spanning the whole target region was not amplified in all cases using our enzyme system and cycling parameters.

Diagnostic conditions were selected by qualitative analysis of PCR yields from set A (WT DNA) and set B (HbS/S DNA). Optimal conditions for zygosity determination are those that amplify each allele-specific amplicon at similar efficiencies, as measured by band intensity from the gel. In this study, the conditions and interactions of reaction 9 (Table 2) satisfied these criteria, producing 517 and 267 bp bands with maximum OD values of 160.67 and 165.93 units, respectively. The optimal nature of these conditions was confirmed unambiguously in a small blind trial of 14 DNA samples of different zygosity (Fig. 3). Reactions generating the 517 bp fragment identified a normal genotype. Amplification of the 267 bp fragment confirmed a homozygous mutation and sickle cell carriers (HbA/S) were identified by the amplification of both the 517 and 267 bp bands, diagnostic of heterozygous inheritance. The robustness of the optimised reaction conditions was demonstrated by the ability to amplify from a range of unknown template concentrations without compromising the allele specificity of the system.

Effect of each parameter on specificity

A quantitative analysis was performed in which the actual product yield for each reaction was used to estimate the effects of individual components on amplification in bi-directional PCR. Signal-to-noise (SNL) ratios were calculated using the maximal optical density (max OD) of the allele-specific PCR products (Table 3) to represent product yield (y) in the quadratic loss function,

$$SNL = -10 \log(1/n\Sigma^{n}_{i=1} 1/y_i^2)$$

assuming that variable y denotes the performance of the system measured for a particular combination of factor settings in a given experiment (total of n repeated measurements per experiment) (17). Reactions giving non-allele-specific fragments or no amplification at all were given a score of 1. The effects of reaction components on the SNL ratio were determined by regression analysis for both normal and HbS/S DNA templates

Table 3. Maximal optical density (max OD) measurements of band intensities from amplification sets A and B in Figure 2

	Reaction								
	1	2	3	4	5	6	7	8	9
Set A, max OD wild-type	5.13	90.83	80.30	63.40	63.70	1	132.27	1	160.67
Set B, max OD mutant	113.6	172.53	169.20	108.73	161.13	183.67	144.80	1	165.93

All values are in arbitrary units.

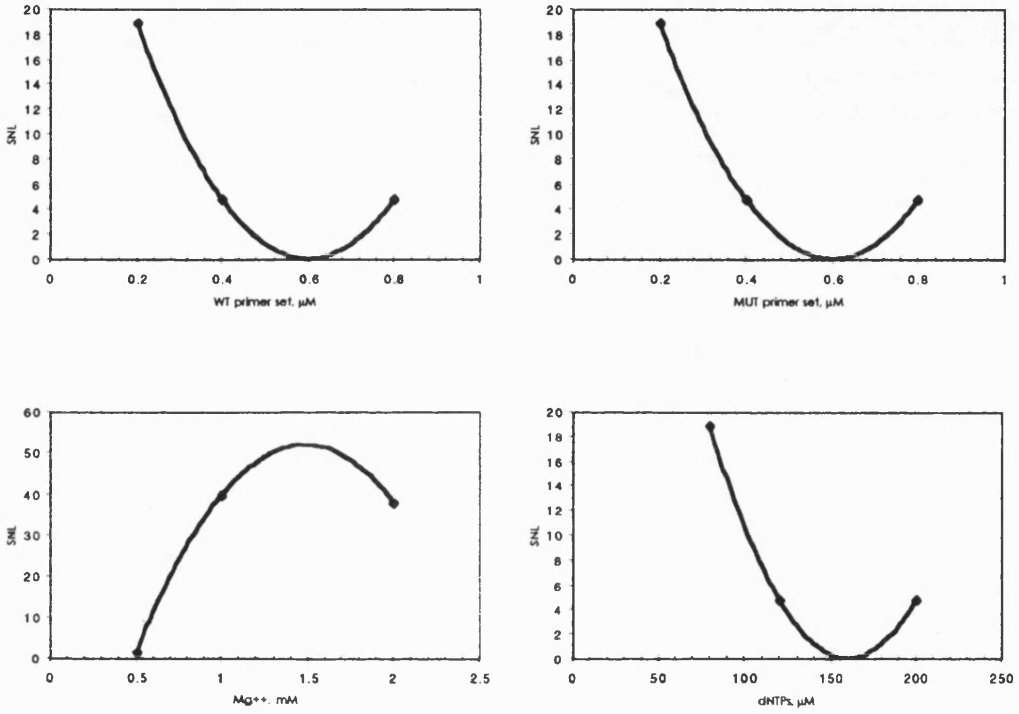


Figure 4. Normal DNA. Effects of reaction components on the SNL ratios for the amplification of a wild-type 517 bp amplicon from normal DNA. Increased SNL value represents a larger effect on product yield than a low SNL.

(Figs 4 and 5, respectively), using calculated SNL values (Table 4).

We were able to demonstrate the significance of priming competition and component interaction on the outcome of bi-directional ASA reactions. The SNL analysis revealed that the optimal magnesium requirement for the generation of maximal product yield without compromising specificity was 1.5 mM. This value did not vary according to template zygosity, which is in accordance with the multiple roles of magnesium in PCR. Likewise, the optimal concentration of dNTPs in sets A and B were similar, with lower levels being favoured. The quantitative binding of magnesium ions by dNTPs may account for this observation, as excess dNTPs may chelate the magnesium cofactor ions of *Taq*, decreasing processivity or simply reducing the yield and specificity.

The regression profiles of primer sets generated by normal and mutant DNA samples were quite different, demonstrating differences in primer stabilities, efficiencies and competition. For normal DNA the aim is to direct amplification of the 517 bp fragment only. The lower concentration of the wild-type primer set tested gave the largest SNL value relating to product yield. Intermediate concentrations decreased the SNL and the upper concentration showed an incremental increase in SNL. An identical profile was observed with the mutant primer set, optimally redundant in the presence of wild-type DNA. It is possible that at lower concentrations of both primer sets, i.e. conditions of minimal competition, the wild-type pair is able to direct allele-specific amplification of the 517 bp fragment.

When the mutant allele is present, the aim is to direct amplification of the 267 bp amplicon only. The optimal concentration of the mutant primer set was predictably at the upper level.

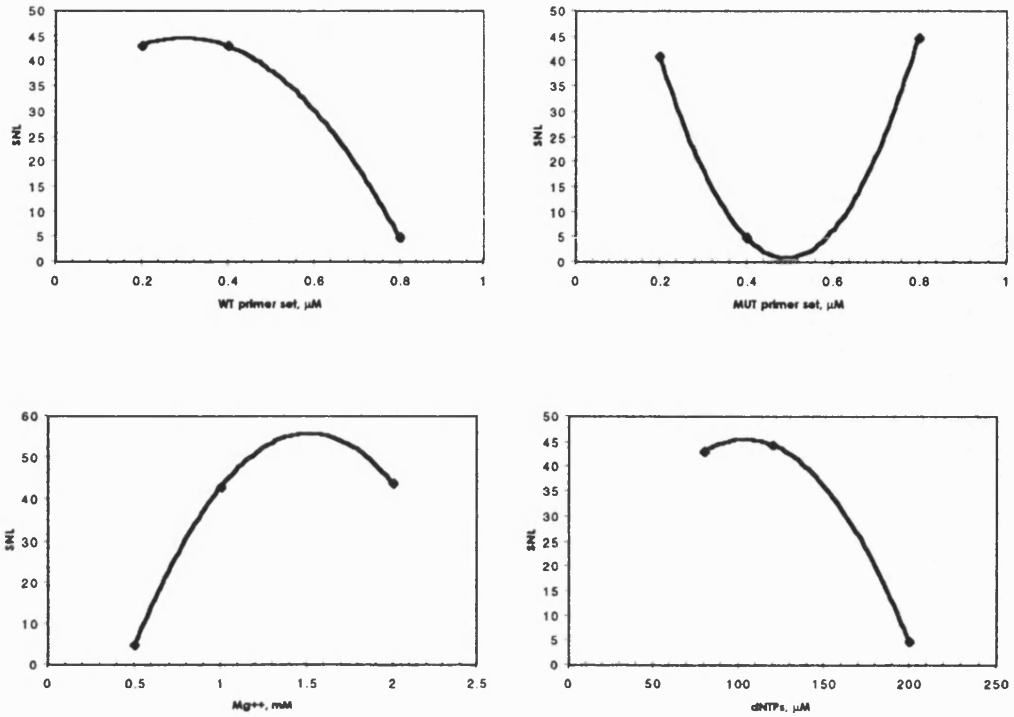


Figure 5. HbS/S DNA. Effects of reaction components on the SNL ratios for the amplification of a mutant 267 bp amplicon from HbS/S DNA. An increased SNL value represents a larger effect on product yield than a low SNL.

Likewise, the inclusion of lower concentrations of the wild-type primer allowed maximal optimal amplification of the mutant fragment, as the wild-type set thereby exhibited least competition with the mutant primer set. We have observed similar trends using other SNP models, confirming that primer interactions have a significant role in determining the outcome of bi-directional amplifications.

DISCUSSION

We have successfully demonstrated the use of bi-directional ASA for single tube genotyping of the SNP responsible for sickle cell anaemia. Our simple matrix optimisation strategy offers advantages over conventional methods, allowing the determination of specific test conditions in a greatly reduced reaction set. Combined with single tube bi-directional allele-specific PCR, rapid diagnosis of normal, homozygous and heterozygous template DNA can be achieved simply and applied to the screening of any known SNP. Bi-directional ASA is based on the recent adaptation of ASA methods, using a single tube for the diagnosis of a two-allele system, differing from other protocols described in which modified oligonucleotides or cycling parameters are generally required to maintain specificity.

We found that the A:A or T:T mismatches designated by the sickle cell polymorphism were completely refractory to amplification under the optimised conditions. This is discordant with a report by Kwok *et al.* (20) in which only A:G, G:A and C:C mismatches reduced overall PCR product yield significantly. Other combinations required the incorporation of additional mismatches near the 3′-end to increase discrimination. This difference may be due to improvements in the process of oligonucleotide synthesis and the use of recombinant *Taq* DNA polymerases. In particular, the use of hot start systems was seen to improve the specificity of our allele-specific amplifications. The optimisation system used and the increased priming competition in our bi-directional assay will have positively contributed to the absolute discrimination.

The implementation of an orthogonal array in our optimisation system generated product yield data from two small reaction sets that served a dual purpose in our experiments. Firstly, we were able to use product yield data qualitatively to identify conditions under which both allele-specific fragments were amplified to similarly maximal efficiencies according to template zygosity. Secondly, reaction 'scores' were used quantitatively in a simple regression analysis, allowing the effects of 'control factors' influencing reaction performance to be assessed simultaneously. This provided an insight into the component interactions and priming competition that dominate bi-directional allele-specific amplifications.

The conditions selected for the blind trial were 0.8 µM for each primer from the wild-type and mutant sets, 1 mM magnesium ions and 80 µM dNTPs. This is unlikely to be in accordance with optimal component levels calculated by SNL values

because of the different objectives of each enquiry. The theoretical target of the SNL analysis is to define conditions producing the maximum yield of each product. This differs from the practical target of specifying diagnostic conditions under which each allele must be amplified specifically and at equal efficiencies enabling a heterozygous mutation to be identified without a gross over-amplification of one allele over another. Thus, to achieve this objective, sub-optimal conditions for the amplification of one allele may be combined with optimal conditions for the amplification of another by a less efficient primer set, as in this study.

Two further variables that have an effect on reaction outcome are the enzyme system and instrumentation used for amplification. The HotStart *Taq* DNA polymerase used in this study was extremely accurate, amplifying allele-specific products with only minimal spurious products. Additionally, the HotStart system was inefficient at amplifying larger fragments (>600 bp) when in competition with a smaller fragment, thus the non-allele-specific fragment spanning the whole target region was not amplified in this case. In similar studies, this larger fragment has been designated an internal control amplification, avoiding the incorporation of a distant primer pair conventionally required for diagnostic tests (12). However, any successful bi-directional ASA reaction will yield a fragment serving as an internal positive control for the PCR. Thus we favour our system where the processivity of *Taq* is expended generating allele-specific products only. We tested a conventional enzyme system in similar experiments and found that the larger fragment was generally produced at the expense of the allele-specific fragment, with decreased specificity (data not shown). Application to other thermal cyclers may require further optimisation, since different reaction vessels and heating/cooling methods can dramatically change the dynamics of a PCR.

Here, genotyping was simply based on size discrimination of the resulting allele-specific PCR products, visualised by agarose gel electrophoresis. With the emergence of qPCR technologies, offering rapid thermal cycling and on-line product detection, our system can be adapted to detect PCR products based on their specific melting temperature using non-specific DNA-binding dyes. This would address high throughput issues whilst maintaining the advantages of low cost and the use of simple chemistries associated with the gel-based assay.

The technique described could be modified to detect more than one SNP simultaneously with the combination of our optimisation strategy and accurate primer design. It has recently been suggested that the grouping and interaction of several SNPs in haplotypes may be more important than single SNPs in the identification of drug response and disease susceptibility of individuals (21). The commercial applications of this are widespread, as the development of reliable, cost-effective tests will be considerably more complex if haplotypes need to be determined rather than single SNPs.

In conclusion, we have described various factors that have a combined influence on the outcome of bi-directional PCR amplifications. We have demonstrated the technique by identifying accurate conditions for determining the zygosity of sickle cell patients and carriers, a technique that can be simply applied to the detection of any known genetic mutation.

Table 4. Signal-to-noise ratios (SNL) for reaction components used for the amplification of (**A**) the 517 bp wild-type allele-specific fragment and (**B**) the 267 bp mutant allele-specific fragment, calculated using product yields from Figure 2

	Level		
SNL	A	B	C
(A) Normal DNA template			
WT-AS/WT-CP517	18.94	4.77	4.77
MUT-AS/MUT-CP267	18.95	4.77	4.77
Mg^{2+}	1.68	39.95	38.12
dNTP	18.94	4.77	4.77
(B) HbS/S DNA template			
WT-AS/WT-CP517	43.12	42.93	4.77
MUT-AS/MUT-CP267	41.01	4.77	44.73
Mg^{2+}	4.77	42.88	43.94
dNTP	42.94	44.32	4.77

ACKNOWLEDGEMENTS

We would like to thank Drs John Old and Chris Fisher at the National Haemoglobinopathy Reference Service, Oxford, UK, for the kind gift of DNA from affected sickle cell individuals and sickle cell carriers. This work was supported by grants from the University of Bath and Molecular Sensing plc.

REFERENCES

1. Brookes,A.J. (1999) The essence of SNPs. *Gene*, **234**, 177–186.
2. Myakishev,M.V., Khripin,Y., Hu,S. and Hamer,D.H. (2001) High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer labelled primers. *Genome Res.*, **11**, 163–169.
3. Lee,H.H. (1996) Ligase chain reaction. *Biologicals*, **24**, 197–199.
4. Bi,W. and Stambrock,P.J. (1997) CCR: a rapid and simple approach for mutation detection. *Nucleic Acids Res.*, **25**, 2949–2951.
5. Nikiforov,T.T., Rendle,R.B., Goelet,P., Rogers,Y.H., Kotewicz,M.L., Anderson,S., Trainor,G.L. and Knapp,M.R. (1994) Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res.*, **22**, 4167–4175.
6. Tonisson,N., Kurg,A., Kaasik,K., Lohmussar,E. and Metspalu,A. (2000) Unravelling genetic data by arrayed primer extension. *Clin. Chem. Lab. Med.*, **38**, 165–170.
7. Gu,W., Aguirre,G.D. and Ray,K. (1998) Detection of single-nucleotide polymorphisms. *Biotechniques*, **24**, 836–837.
8. Kuklin,A., Munson,K., Gjerde,D., Haefele,R. and Taylor,P. (1997/98) Detection of single-nucleotide polymorphisms with the WAVE™ DNA fragment analysis system. *Genet. Test.*, **1**, 201–205.
9. Newton,C.R., Graham,A., Heptinstall,L.E., Powell,S.J., Summers,C., Kalsheker,N., Smith,J.C. and Markham,A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.*, **17**, 2503–2516.
10. Gibbs,R.A., Nguyen,P. and Caskey,C.T. (1989) Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Res.*, **17**, 2437–2448.
11. Ye,S., Humphries,S., and Green,F. (1992) Allele specific amplification by tetra-primer PCR. *Nucleic Acids Res.*, **20**, 1152.
12. Liu,Q., Thorland,E.C., Heit,J.A. and Sommer,S.S. (1997) Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. *Genome Res.*, **7**, 389–398.
13. Sasvari-Szekely,M., Gerstner,A., Ronai,Z., Staub,M. and Guttman,A. (2000) Rapid genotyping of factor V Leiden mutation using single-tube

bi-directional allele-specific amplification and automated ultrathin-layer agarose gel electrophoresis. *Electrophoresis*, **21**, 816–821.
14. Sommer,S.S., Groszbach,A.R. and Bottema,C.D.K. (1992) PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques*, **12**, 1152.
15. Wu,Y., Ugozzoli,L., Pal,B.K. and Wallace,R.B. (1989) Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl Acad. Sci. USA*, **86**, 2757–2760.
16. Old,J., Petrou,M., Varnavides,L., Layton,M. and Modell,B. (2000) Accuracy of prenatal diagnosis for haemoglobin disorders in the UK: 25 years' experience. *Prenat. Diagn.* **20**, 986–991.
17. Cobb,B.D. and Clarkson,J.M. (1994) A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. *Nucleic Acids Res.*, **22**, 3801–3805.
18. Henegariu,O., Heerema,N.A., Dlouhy,S.R., Vance,G.H. and Vogt,P.H. (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23**, 504–511.
19. Linz,U., Delling,U. and Rubsamen-Waigmann,H. (1990) Systematic studies on parameters influencing the performance of the polymerase chain reaction. *J. Clin. Chem. Clin. Biochem.*, **28**, 5–13.
20. Kwok,S., Kellogg,D.E., McKinney,N., Spasic,D., Goda,L., Levenson,C. and Sninsky,J.J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.*, **18**, 999–1005.
21. Davidson,S. (2000) Research suggests importance of haplotypes over SNPs. *Nature Biotechnol.*, **18**, 1134–1135.

Appendix IV

Rapid SNP genotyping using Fluorescent Bi-directional PCR

Short Technical Reports

SNP Genotyping Using Single tube SNP Fluorescent Bidirectional PCR

BioTechniques 33:__-__ (July 2002)

ABSTRACT

SNP genotyping is a well-populated field with a large number of assay formats offering accurate allelic discrimination. However, there remains a discord between the ultimate goal of rapid, inexpensive assays that do not require complex design considerations and involved optimization strategies. We describe the first integration of bidirectional allele-specific amplification, SYBR Green® I, and rapid-cycle PCR to provide a homogeneous SNP-typing assay. Wild-type, mutant, and heterozygous alleles were easily discriminated in a single tube using melt curve profiling of PCR products alone. We demonstrate the effectiveness and reliability of this assay with a blinded trial using clinical samples from individuals with sickle cell anemia, sickle cell trait, or unaffected individuals. The tests were completed in less than 30 min without expensive fluorogenic probes, prohibiting design rules, or lengthy downstream processing for product analysis.

INTRODUCTION

Although PCR was described nearly two decades ago (11), it still remains a core enabling technology for many diagnostic procedures. More recently, there has been renewed interest in the use of PCR for detecting SNPs since it has the potential to offer rapid and inexpensive diagnosis. SNP genotyping has witnessed a progression from platforms requiring secondary analysis of PCR products [e.g., RFLP (2), single-strand conformational polymorphism (8), and DNA sequencing (14)] to rapid homogeneous genotyping with sequence-specific detection (10,20). This evolution has largely been accomplished through the availability of PCR instruments that combine amplification and product detection. Continued improvement of instrumentation in terms of speed and detection sensitivity suggests that PCR may be an effective platform on which to base SNP-typing assays for diagnostic application.

To this end, a number of different fluorescent chemistries have been developed that provide sequence-specific probes for real-time PCR. These typically employ fluorescence resonance energy transfer (FRET) between two fluorophores, with one quenching the fluorescence of the other (3,17,21,22). The design, synthesis, and purification of these probes are both difficult and expensive, which limits their commercial suitability.

A widely used method for mutation detection that does not use fluorescent probes is allele-specific amplification (ASA) (12,15) and is based on the observation that a mismatch between template and PCR primer reduces or prevents amplification by *Taq* DNA polymerase. This process is assisted by the lack of a 3′→5′ proofreading activity in the enzyme. Two allele-specific primers complementary to the wild-type and mutant sequence are used to amplify target DNA in a pairwise manner with a single common primer. The presence or absence of PCR product in each reaction mixture indicates whether the normal or mutant allele is present in the target. Typically, two reactions are required to genotype a two-allele system, and internal control amplification is necessary to provide a positive control for the PCR test (19).

More recently, we introduced a single-tube adaptation called bidirectional ASA (23), in which two allele-specific primer pairs are included in a single PCR test. Like similar methods published independently [e.g., PCR-CTPP (5), SB-ASA (6), bidirectional AS-PCR (7), bi-PASA (9), SAS-PCR (16), tetra primer PCR (27), and tetra primer ARMS PCR (28)], this strategy has benefits in terms of reducing reagents and consumable costs, yet requires a secondary analysis for product identification by gel electrophoresis or chromatography, which decreases overall throughput.

We describe a new assay format called fluorescent bidirectional PCR, which represents the first integration of bidirectional PCR with detection of specific PCR products according to their T_m. This introduces homogeneity to a previously heterogeneous system and takes advantage of contemporary PCR instruments, which offer increased speed and cost-effectiveness. The test is compatible with rapid-cycle PCR, decreasing the time taken for bidirectional PCR methods from 4 h to 35 min. In addition, specificity is improved since shorter extension times strongly disfavor mismatched primer extension (17,25). The use of a single generic DNA binding dye, in this instance SYBR Green® I, has enabled inexpensive characterization of the amplification profile, which offers an interesting, more rapid alternative to the T_m-shift genotyping assay previously described (4).

Critically, fluorescent bidirectional PCR does not require any additional primer modifications such as GC tails (4) or additional mismatches near the 3′ end of the primers to improve specifici-

Table 1. Primer Titration Matrix

Reaction	SCF-WT	WT-CP 86	SCF-MUT	MUT-CP112
1	A	A	A	A
2	A	A	B	B
3	A	A	C	C
4	B	B	A	A
5	B	B	B	B
6	B	B	C	C
7	C	C	A	A
8	C	C	B	B
9	C	C	C	C

Primers were aliquoted according to this matrix, where A = 0.15 μM, B = 0.3 μM, and C = 0.9 μM. Inter-set primer concentrations were chosen empirically to cover a wide range.

ty and utilizes a commercially available reaction mixture with no additional reagents to those normally employed for a standard PCR assay. Since the true cost of a genotyping test is defined by not only reagent and consumable costs but also time and labor for design and optimization, our simplified strategy adds to overall cost-effectiveness. To optimize any new PCR effectively, a magnesium and oligonucleotide titration is usually performed, and once conditions have been established for a particular SNP, re-optimization is unnecessary.

We demonstrate the efficiency and reliability of the assay by detecting the single-base-pair mutation that causes the autosomal recessive disease, sickle cell anemia. Here, a single A→T transversion in the sequence encoding codon 6 of the human β-globin gene expresses itself as an amino acid substitution of glutamine to valine, forming a mutant globin chain (HbS) (26). In this study, a set of clinical DNA samples was genotyped to validate fluorescent bidirectional PCR as an interesting and cost-effective alternative to other SNP genotyping methodologies.

MATERIALS AND METHODS

Fluorescent Bidirectional PCR Primers

Oligonucleotides were designed using PrimerCalc© software (Q-Biogene, CA, USA) and synthesized by Sigma-Genosys, with a reverse-phase cartridge purification step. Each primer was designed with a T_m of 60°C to allow generic PCR cycling conditions. SCF-WT (5′-ATGGTGCACCTGACTCCTGA-3′) and WT-CP86 (5′-AGGGCCTCACCACCAACTTC-3′) direct the amplification of an 86-bp product from the normal β-globin gene, with a calculated T_m of 87°C (wild-type primer set). SCF-MUT (5′-GCAGTAACGGCAGACTTCTCCA-3′) and MUT-CP112 (5′-AGGGCAGAGCCATCTATTGC-3′) amplify a 112-bp fragment from homozygous mutant DNA, with a calculated T_m of 85°C (mutant primer set). Product melting temperatures were predicted using PrimerCalc software, which employs nearest neighbor thermodynamic algorithms for T_m calculations (1).

Fluorescent Bidirectional PCR Assay

Reactions were performed in a total volume of 20 µL, containing 2 µL LightCycler™ DNA FastStart SYBR Green I mixture (*Taq* DNA polymerase, reaction buffer, deoxynucleoside triphosphate mixture, and SYBR Green I

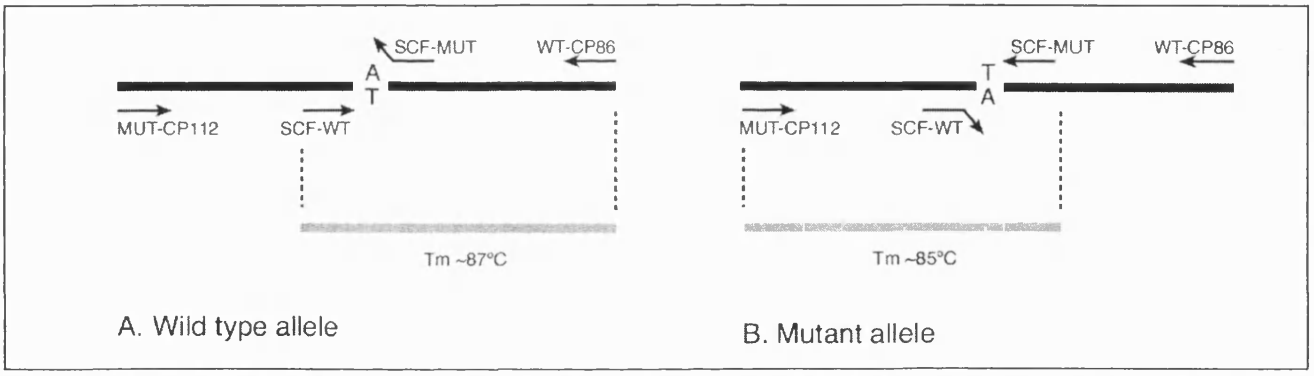


Figure 1. Schematic of fluorescent bidirectional PCR. In a single reaction with two allele-specific primer sets, genomic DNA (■) will be amplified by either set depending on template zygosity, yielding allele-specific PCR products (■). Straight arrows indicate complementarities between primer and template, whereas upward-/downward-facing arrows represent a single-base mismatch at the 3′ end of the allele-specific primer. (A) Samples homozygous for the wild-type allele will be amplified by SCF-WT and WT-CP86, giving a product with a predefined T_m of 87°C. (B) Samples homozygous for the mutant allele will generate a product with a predefined T_m of 85°C. Heterozygous inheritance will amplify both allele-specific PCR products. Internal control primers are not necessary since at least one allele-specific fragment will be amplified per reaction under optimized conditions.

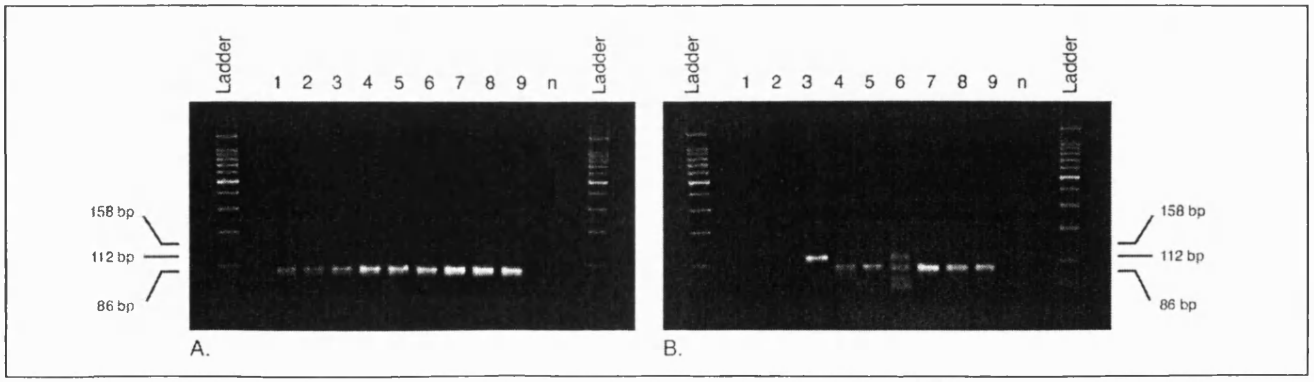


Figure 2. Gel images from primer titrations. Reactions 1-9 correspond to conditions described in Table 1 using (A) wild-type and (B) mutant human genomic DNA template. Samples void of template are denoted by "n". Actual product T_m values for the 86-bp product were 87.1°C (±0.3°C), and 85.7°C for the 112-bp fragment. The 158-bp target region was not amplified to a detectable level across the reaction set.

dye; Roche Applied Science, Lewes, UK), a final concentration of 3 mM $MgCl_2$ and 20 ng human genomic DNA. For primer optimization sets, oligonucleotides were aliquoted according to Table 1, with control DNA templates. For the blinded trial, optimal primer ratios, as defined by reaction 3, Table 1, were used (0.15 µM each wild-type Primer, and 0.9 µM each mutant primer) with 45 DNA samples of unknown genotype.

PCR Cycling and Melt Curve Analysis

PCR cycling and melt analysis protocols were performed on the LightCycler PCR instrument (Roche Applied Science). Amplifications were completed in approximately 30 min, including a 10-min *Taq* DNA polymerase activation step at 95°C, 35 PCR cycles of denaturation at 95°C for 1 s, primer annealing at 62°C for 3 s, and primer extension for 0 s at 72°C. All temperature transition rates were set to 20°C/s and a single fluorescence acquisition was made at 72°C for each cycle. Fluorescence gains (used to set sensitivity of detection) for the LightCycler were F1-2 throughout.

Product identification and specificity was confirmed by executing a melting curve analysis immediately after amplification in the same closed capillary tube. An initial denaturation at 95°C for 0 s was followed by a cooling step to 67°C (5°C above primer annealing, ensuring all fluorescence from dsDNA originates from product-to-product re-annealing) and a slow denaturation phase to 95°C at a rate of 0.1°C/s, with continual fluorescence acquisition. Product denaturation was observed as a rapid loss of fluorescence near the calculated T_m. This data was converted into melting peaks by LightCycler software, plotting the first negative derivative of fluorescence loss with respect to temperature (-d(F1)/dT vs. T). The apex of the melt peak represents product T_m.

Genotype Validation by Restriction Digestion

To validate genotype scoring by fluorescent bidirectional PCR, conventional PCR was performed using the DNA samples of unknown genotype, followed by restriction enzyme digestion with *Dde*I. Each PCR was a total volume of 25 µL, containing 20 ng template DNA, 0.5 µM each outer primer (WT-CP86 and MUT-CP112), and 10 µL PCR Master Mix (Promega) containing *Taq* DNA polymerase, dNTPs, $MgCl_2$, and reaction buffers at optimal concentrations for efficient amplification. All reactions were performed using an MJ Thermal Cycler (GRI, UK) through an initial denaturation at 95°C for 2 min, followed by 30 cycles of denaturation at 95°C for 30 s, primer annealing at 60°C for 30 s, and extension at 72°C for 35 s. A 10-µL aliquot of the reaction was mixed with 20 U *Dde*I (Promega) and 3 µL One-Phor-All *Plus* buffer (Amersham Pharmacia, Little Chalfont, UK), made up to 30 µL with PCR-grade water. The solutions were incubated at 37°C for 1 h, followed by an enzyme inactivation step at 65°C for 15 min. Digestion products were separated according to size on a 2.5% agarose gel (Q-Biogene) and visualized by staining with ethidium bromide. This protocol was used for all gel-based separations.

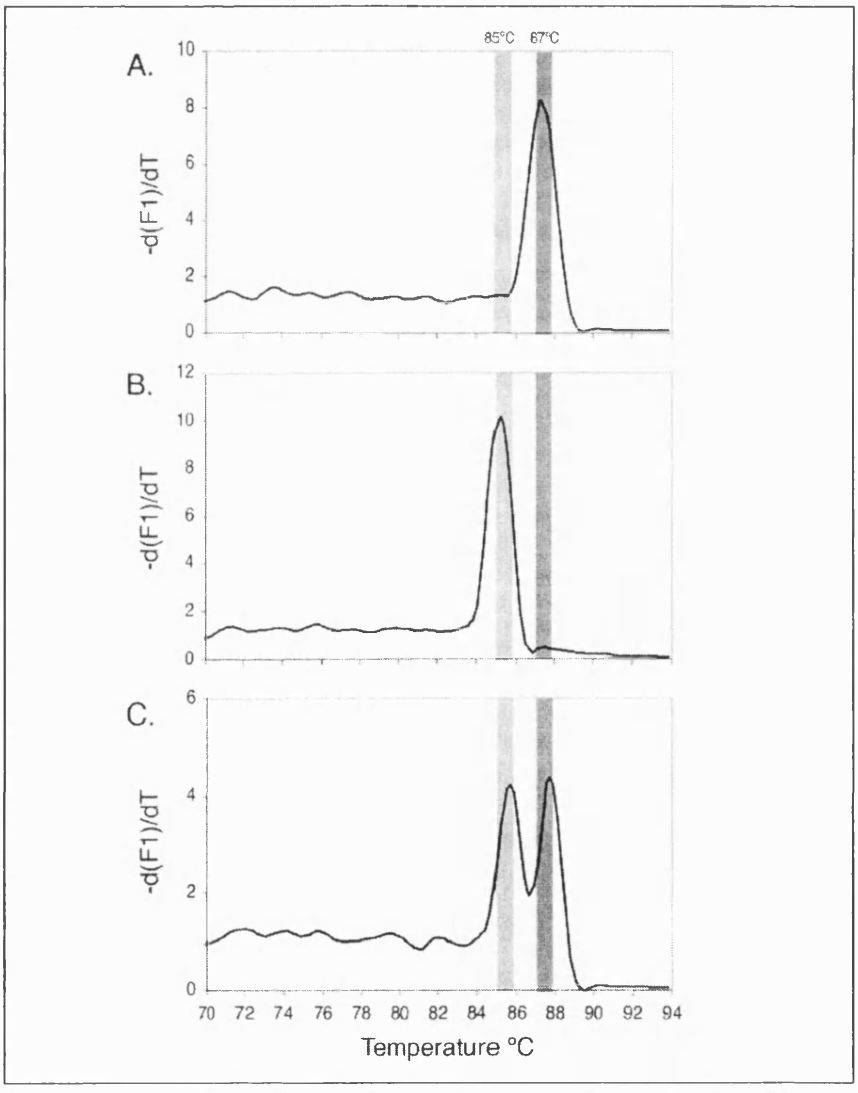


Figure 3. Example melting profiles from diagnostic tests representing one of each possible genotype. (A) Homozygous wild-type (sample 4); (B) homozygous mutant (sample 6); and (C) heterozygous DNA (sample 10). The three genotypes are easily identified by visual inspection of the graphs as described in text. Actual T_m values were slightly higher than calculated in most cases.

RESULTS AND DISCUSSION

Assay Design

The principle of the fluorescent bidirectional PCR assay is shown schematically in Figure 1. A key requirement of the test is the ability to design appropriate primer pairs that amplify allele-specific PCR products with a predicted T_m difference of 2°C, allowing adequate discrimination by melt curve analysis. This can be accomplished by defining primer T_m value and searching for outer primers that generate products between 50 and 150 bp. Smaller products are recommended, giving the widest variation in melt temperature, as well as being the size of choice for rapid cycling protocols (18,24). The thermal profile employed in our experiments can be transferred between assays following these guidelines, and a slower, more-controlled temperature transition during melting may enhance T_m discrimination further (13).

We chose the sequence surrounding the sickle cell mutation site to demonstrate the applicability of the test, aided by the high GC content downstream of the SNP site that formed the wild-type fragment (58%) compared to the upstream sequence that formed mutant fragment (49%). As with all genotyping assays, the actual SNP position and surrounding sequence is nonvariable, meaning that under rare conditions the assay may not be appropriate. However, using the described procedure, we have successfully designed primer sets to detect a range of sequence polymorphisms using conventional primer design software.

Cycling conditions were manipulated to favor smaller products by reducing the extension time to 0 s, relying on the transitional times of the kinetic PCR cycle to amplify the shorter products (16). *Taq* DNA polymerase extends the two primers flanking the target region regardless of genotype, generating a 158-bp product with a calculated T_m of approximately 88°C. The melt peak from this larger fragment would be indistinguishable from the wild-type product, which has a slightly lower T_m of 87°C. Thus, amplification of this fragment may lead to false results by masking ASAs. Very rapid cycling combined with internal competition from smaller fragments and a high-fidelity enzyme system totally suppressed amplification of this region throughout our experiments (Figure 2).

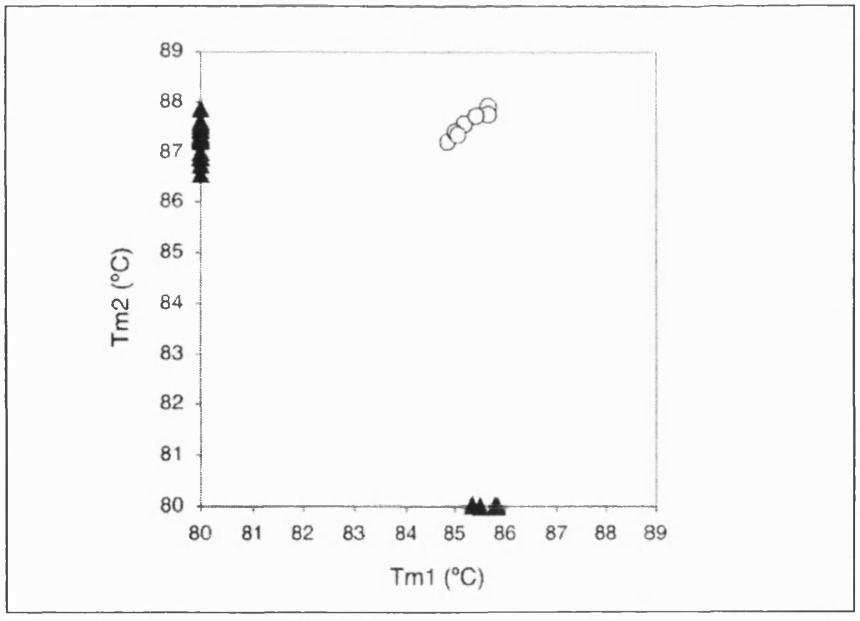


Figure 4. Cluster analysis to determine sickle cell genotype. The x-axis represents peaks within the actual T_m range of 85.4°C (±0.3°C), and the y-axis represents areas within the actual T_m range of 87.3°C (±0.3°C). A symbol is assigned to represent homozygous DNA (▲) or heterozygous DNA (○). Points lining up along either the x- or y-axis diagnose homozygous inheritance of mutant or wild-type alleles, respectively, and points lying near to the x = y axes diagnoses heterozygous inheritance of the sickle cell mutation.

Assay Optimization

We specify optimal discriminatory conditions as those able to amplify allele-specific products *only* with similar efficiency as determined by melt peak height and peak area (13). In addition, these conditions should reduce or completely remove artifactual amplification in the form of primer dimers. We used a strategy that would normally be applied to new PCR optimization, modified to accommodate oligonucleotide pairs. An initial magnesium titration was performed using "outer" primers to direct amplification. Maximum specificity and product yield was achieved with 3 mM magnesium (data not shown). Two oligonucleotide titration sets using control DNAs of known genotype under rapid cycling conditions defined optimal primer set concentrations for allelic discrimination. The ratios and interactions of reaction 3 gave optimal results, yielding single products with a T_m of 87.4°C with wild-type DNA and 85.7°C with mutant DNA. Each melt peak had a height of approximately 10 fluorescence units and similar peak areas. The size of corresponding products was confirmed by gel electrophoresis (Figure 2).

SNP Genotyping

The fluorescent bidirectional PCR assay was validated using optimized discriminatory conditions to genotype 45 unknown samples for sickle cell status based on their melting profiles alone. In each case, clearly distinct profiles were achieved depending on genotype (Figure 3). Thirty-three samples generated a melt peak around 87.5°C (±0.2°C), five samples produced a melt peak at 85.7°C (±0.2°C), and seven samples amplified both allele-specific fragments, giving a dual-peak with maxima at 85.3°C (±0.3°C) and 87.6°C (±0.3°C), thus diagnosing homozygous wild-type, homozygous mutant, and heterozygous inheritance, respectively. Cluster analysis of product T_m values showed three unambiguous and distinct clustering regions, each corresponding to a particular genotype (Figure 4). For

each test, "no template control" samples were included to assess contamination and nonspecific amplification by primer sets; no amplification occurred in these controls. If primer dimer had been formed in the reaction, then the T_m of the artifact would ensure that points were well away from profiles generated by actual template amplifications.

Experienced users of melt curve analysis will acknowledge that there can be well-to-well variation in the actual T_m of PCR products. Such variation is usually attributed to SYBR Green I concentration, salt concentration, or actual product yield (13). By using a commercial master mixture with a consistent concentration of SYBR Green I and employing optimized magnesium concentrations, this variation was minimized. In practice, the deviations observed were likely due to differences in product yield and did not cause error in our tests. Melt peak data was 100% concordant with restriction digestion band patterns (data not shown), validating our assay entirely.

In summary, fluorescent bidirectional PCR is mediated through a highly efficient kinetic process, coupled with an accurate melt curve analysis. The data presented here suggest that it has potential as a tool for screening programs of sickle cell anemia, particularly if coupled with informatics tools that would allow automation. While we appreciate that for clinical use a more thorough evaluation of the test would be appropriate, our trial has adequately demonstrated the potential of this method for use in SNP research. It provides a rapid, single-tube format, with internal control and sequence validation for SNP screening, in particular for near patient testing where the cost-effectiveness of the assay is the main driver. We believe that this will prompt a revival in an interest in allele-specific PCR for SNP detection.

REFERENCES

1. **Breslauer, K.J., R. Frank, H. Blocker, and L.A. Marky.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA *83*:3746-3750.
2. **Chang, J.C. and W.K. Kan.** 1982. A sensitive new prenatal test for sickle-cell anemia. N. Engl. J. Med. *307*:30-36.
3. **Crockett, A.O. and C.T. Wittwer.** 2001. Fluorescein-labeled oligonucleotides for real-time PCR: using the inherent quenching of deoxyguanosine nucleotides. Anal. Biochem. *290*:89-97.
4. **Germer, S. and R. Higuchi.** 1999. Single-tube genotyping without oligonucleotide probes. Genome Res. *9*:72-78.
5. **Hamajima, N.** 2001. PCR-CTPP: a new genotyping technique in the era of genetic epidemiology. Exp. Rev. Mol. Diagn. *1*:119-123.
6. **Jiang, Z., J. Shi, S. Yang, C. Zhang, H. Jiang, Z. Chen, L. Jin, D. Lu, and W. Huang.** 2001. A simple and rapid new method for SNP typing by single-tube bidirectional allele-specific amplification. Zhonghua Yi Xue Yi Chuan Xue Za Zhi *18*:306-309.
7. **Karhukorpi, J. and R. Karttunen.** 2001. Genotyping interleukin-10 high and low producers with single-tube bidirectional allele-specific amplification. Exp. Clin. Immunogenet. *18*:67-70.
8. **Lareu, R.R., N.R. Swanson, and S.A. Fox.** 1997. Rapid and sensitive genotyping of hepatitis C virus by single strand conformation polymorphism. J. Virol. Methods *64*:11-18.
9. **Liu, Q., E.C. Thorland, J.A. Heit, and S.S. Sommer.** 1997. Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. Genome Res. *7*:389-398.
10. **Livak, K.J.** 1999. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet. Anal. Biomol. Eng. *14*:143-149.
11. **Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich.** 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold Spring Harbor Simple Quantitative Biol. *51*:263-273.
12. **Newton, C.R., A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, and A.F. Markham.** 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. *17*:2503-2516.
13. **Ririe, K., R.P. Rasmussen, and C.T. Wittwer.** 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. *245*:154-160.
14. **Ronaghi, M., S. Karamohamed, B. Petterson, M. Uhlen, and P. Nyren.** 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. *242*:84-89.
15. **Sarkar, G., J. Cassady, C.D.K. Bottema, and S.S. Sommer.** 1990. Characterisation of polymerase chain reaction amplification of specific alleles. Anal. Biochem. *186*:64-68.
16. **Sasvari-Szekely, M., A. Gerstner, Z. Ronai, M. Staub, and A. Guttman.** 2000. Rapid genotyping of factor V Leiden mutation using single-tube bidirectional allele-specific amplification and automated ultrathin-layer agarose gel electrophoresis. Electrophoresis *21*:816-821.
17. **Silva, D.D., A. Reiser, M. Herrmann, K. Tabiti, and C.T. Wittwer.** 1998. Rapid Genotyping and quantification on the LightCycler™ with hybridization probes. Biochemica 2:12-15.
18. **Silva, D. and C.T. Wittwer.** 2000. Monitoring hybridisation during polymerase chain reaction. J. Chromatogr. B. Biomed. Appl. *741*:3-13.
19. **Sommer, S.S., A.R. Groszbach, and C.D.K. Bottema.** 1992. PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. BioTechniques *12*:82-87.
20. **Tapp, I., L. Malmberg, E. Rennel, M. Wik, and A.C. Syvanen.** 2000. Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5′-nuclease TaqMan assay and molecular beacon probes. BioTechniques *28*:732-738.
21. **Thelwell, N., S. Millington, A. Solinas, J. Booth, and T. Brown.** 2000. Mode of action and application of Scorpion primers to mutation detection. Nucleic Acids Res. *28*:3752-3761.
22. **Tyagi, S., S.A.E. Marras, and F.R. Kramer.** 2000. Wavelength-shifting molecular beacons. Nat. Biotechnol. 18:1191-1196.
23. **Waterfall, C.M. and B.D. Cobb.** 2001. Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. Nucleic Acids Res. *29*:e119.
24. **Wittwer, C.T. and D.J. Garling.** 1991. Rapid cycle DNA amplification: time and temperature optimization. BioTechniques *10*:76-83.
25. **Wittwer, C.T., B.C. Marshall, G.H. Reed, and J.L. Cherry.** 1993. Rapid cycle allele-specific amplification: studies with the cystic fibrosis ΔF508 locus. Clin. Chem. *39*:804-809.
26. **Wu, Y., L. Ugozzoli, B.K. Pal, and R.B. Wallace.** 1989. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA *86*:2757-2760.
27. **Ye, S., S. Dhillon, X. Ke, A.R. Collins, and I.N.M. Day.** 2001. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. *29*:e88.
28. **Ye, S., S. Humphries, and F. Green.** 1992. Allele specific amplification by tetra-primer PCR. Nucleic Acids Res. *20*:1152.

We thank Drs. John Old and Chris Fisher of the National Haemoglobinopathies Reference Service (Oxford, UK) for genotyping controls and unknown DNA samples. This work was supported by research grants from Molecular Sensing plc and the University of Bath. *Address correspondence to Dr. Christy M. Waterfall, Molecular Sensing plc, Challeymead Business Park, Bradford Road, Melksham, Wiltshire SN12 8LH, UK. e-mail: christy.waterfall@molecular-sensing.com.*

Received 21 February 2002; accepted 27 March 2002.

Christy M. Waterfall and Benjamin D. Cobb
Molecular Sensing
Melksham, Wiltshire, UK

Appendix V

Locked Nucleic Acids as enhancers of allele-specific PCR

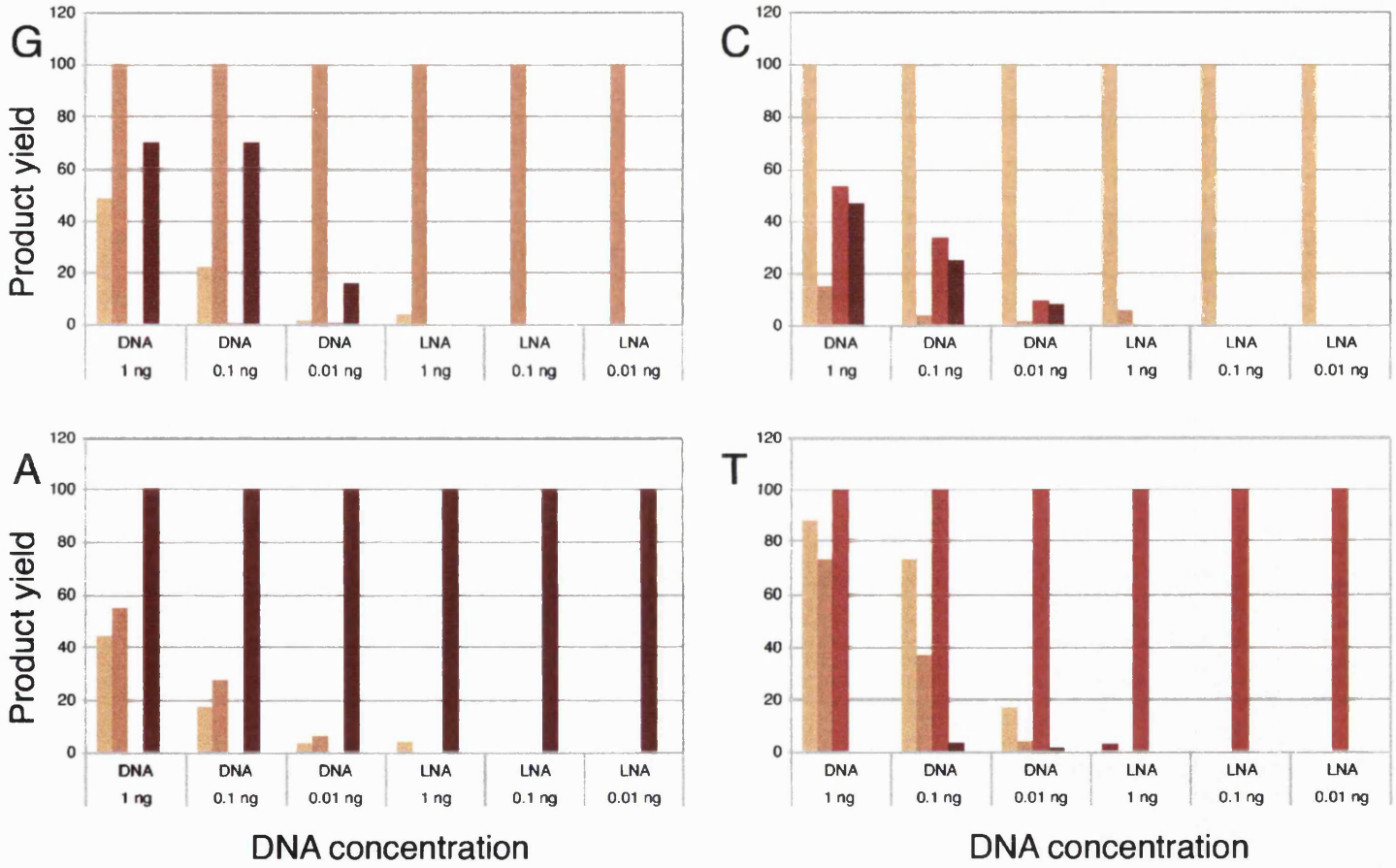
This brief report summarises experiments comparing the allele-specificity of PCR amplification directed by conventional DNA oligonucleotides or LNA TrueSNP™ primers (Proligo LLC, Boulder, USA).

Assay design

Four plasmid DNA templates were employed for the PCR tests, each identical in sequence apart from a single variant nucleotide (G, C, A or T) at position 11 of a cloned β-actin gene insert. To assess specificity across a range of starting template concentrations, three dilutions were tested in each case (1 ng, 0.1 ng and 0.01 ng). Four allele-specific DNA primers (WT_AS_X) and a universal common primer (WT_OP_98) directed the amplification of a 98 bp amplicon with a Tm of 85 °C. For LNA experiments, equivalent TrueSNP™ primers replaced allele-specific DNA primers. Previously optimised reagent concentrations and cycling protocol were used for PCR amplifications on the LightCycler™. Melt peaks were obtained at the end of each cycle, and peak area was used to represent product yield. Each experiment was run in triplicate, and is displayed as an average across the set.

Results and Summary

These data represent amplifications primed by the 4 possible matched primer/templates and 12 mismatched primer/templates, over a range of DNA concentrations. The data clearly shows a significant improvement in specificity using LNA primers compared to DNA primers at all template concentrations. When DNA oligos were used, mismatched amplifications were apparent in nearly all cases. LNA primers rarely extended a mismatched primer, and then only at the highest starting template concentration, demonstrating their effectiveness in improving allele-specific PCR methods.



Amplifications primed by allele-specific primer with a 3' terminal 'C' is shown in gold, 'G' is yellow, 'A' is red, and 'T' is dark red.

Appendix VI

Kinetic characterisation of primer mismatches in allele-specific PCR: A quantitative assessment

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 299 (2002) 715–722
BBRC
www.academicpress.com

Kinetic characterisation of primer mismatches in allele-specific PCR: a quantitative assessment

Christy M. Waterfall,[a,b,*] Robert Eisenthal,[b] and Benjamin D. Cobb[a]

[a] *Molecular Sensing plc, Challeymead Business Park, Bradford Road, Melksham, Wiltshire SN12 8LH, UK*
[b] *Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7RY, UK*

Received 5 November 2002

Abstract

A novel method of estimating the kinetic parameters of *Taq* DNA polymerase during rapid cycle PCR is presented. A model was constructed using a simplified sigmoid function to represent substrate accumulation during PCR in combination with the general equation describing high substrate inhibition for Michaelis–Menten enzymes. The PCR progress curve was viewed as a series of independent reactions where initial rates were accurately measured for each cycle. Kinetic parameters were obtained for allele-specific PCR (AS-PCR) amplification to examine the effect of mismatches on amplification. A high degree of correlation was obtained providing evidence of substrate inhibition as a major cause of the plateau phase that occurs in the later cycles of PCR.

Keywords: Kinetic model; Allele-specific PCR; Michaelis–Menten; Substrate inhibition; Sigmoid function

The polymerase chain reaction (PCR) enables replication of nucleic acid in vitro through the action of a thermally stable DNA polymerase [1,2]. Until recently it has been used predominantly as a binary tool to indicate the presence or absence of a particular target sequence. Instruments capable of quantitative detection of amplification in real time (qPCR) have enabled sensitive analysis of gene expression and accurate determination of copy number [3,4], enhancing the utility of PCR as a diagnostic tool.

Despite the routine use of PCR in molecular biology, enzymatic and kinetic studies have been limited. Early characterisation studies detailed the efficiency with which *Taq* DNA polymerase differentiates between matched and mismatched 3′ nucleotides [5–15]. These analyses used end-point methods that considered the action of *Taq* in isolation of PCR dynamics. This significantly underestimates the kinetic effects known to influence reaction fidelity and efficiency, and limits application in designing genotyping tests. Furthermore, the rapid cycling of contemporary qPCR instruments has been shown to strongly disfavour mismatch extension [16], suggesting that abrupt transitions in temperature have effects that have not been previously considered.

Several theoretical descriptions of PCR have been put forward that apply different mathematical approaches to simulate various physical parameters of the system. These include statistical estimations of amplification rate [17], probability of DNA replication [18], probability of DNA binding rates [19], and derivation of expressions for amplification efficiency [20–23]. Each model has yet to be evaluated using real PCR data.

Michaelis–Menten kinetics have been used to describe the variation of the rates of many enzyme-catalysed reactions as the substrate or effector concentration is varied. There is experimental evidence to show that DNA polymerase reactions obey this typical behaviour [24]. Therefore, the Michaelis–Menten system can be utilised to underlie the principles of a theoretical description and obtain quantitative predictions of enzyme characteristics including catalytic activity, effects of specific inhibitors, and an enzyme's specificity toward a particular substrate [25].

* Corresponding author. Fax: +44-0-1225-700-056.
E-mail address: christy.waterfall@molecular-sensing.com (C.M. Waterfall).

0006-291X/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0006-291X(02)02750-X

We describe a mathematical model based on the Michaelis–Menten framework that enables the kinetic characterisation of PCR. In particular, we apply this model to rapid cycle allele-specific PCR (AS-PCR) that is used to discriminate between alleles of a gene based on single base pair differences. AS-PCR relies on the ability of *Taq* DNA polymerase to extend a primer only when its 3′ end is perfectly complementary to the template. The allele of the target DNA can be inferred by the presence or absence of amplicon [12]. These data would be directly applicable to designing single nucleotide polymorphism (SNP) detection experiments for disease and pharmacogenomic analysis [26].

In our experiments, PCR was shown to follow the general equation for high substrate inhibition. In this formulation an additional term is introduced into the denominator of the simple Michaelis–Menten equation; this term becomes predominant as the substrate concentration increases, resulting in inhibition of the observed rate [27]. This formed the foundation of our model, which was used to represent PCR kinetically using a single equation that considers each cycle of PCR as a defined enzymatic reaction. A simple sigmoidal function was used to represent changing substrate concentration over the course of the reaction. A high degree of correlation was obtained, suggesting that PCR is sensitive to substrate during the plateau phase of the amplification.

Materials and methods

Plasmid DNA model system. All oligonucleotides were designed using PrimerCalc software (Q-Biogene; Carlsbad, USA) and synthesised by Sigma–Genosys (Cambridgeshire, UK) with a reverse-phase cartridge purification step. A 374 bp insert from the β-actin gene was amplified from human genomic DNA (Sigma–Aldrich; Dorset, UK) directed by primers BF (5′-TTC CGT AGG ACT CTC TTC TCT GA-3′) and BR2 (5′-GGG GTG TTG AAG GTC TCA AAC AT-3′). The resulting amplicon was purified using Wizard PCR Purification Kit (Promega; Madison, WI). Mutagenic primers, BFR (5′-CGG GAT CC TTC CGT AGG ACT CTC TTC TCT GA-3′) and BR2X (5′-CGG GAT CC XGA GGG GTG TTG AAG GTC TCA AAC AT-3′ where X is G, C, A, and T) were used to re-amplify the purified PCR product introducing a 5′ tail containing a variant nucleotide. The re-amplified products were ligated into plasmid DNA using a pGEM-T Easy Vector System I Kit (Promega), following manufacturer's protocol. Competent *Escherichia coli* cells (Q-Biogene) were transformed with the ligation mixes and resulting clones were screened colorimetrically to confirm transformation. Plasmid Minipreps (Qiagen; Sussex, UK) were used to purify the plasmid DNA following manufacturer's protocol. Plasmid preparations were sequenced to confirm the presence and position of the variant nucleotide. Preparations were quantified by measuring absorbance at 260 nm using a Cary-100 UV–Visible Spectrophotometer (Varian; Surrey, UK). Plasmid templates were used to model a "susceptibility gene" covering all base configurations.

Allele-specific PCR amplifications. PMOP98 (5′-GCTGTCCCCAG TGGCTT-3′) and PMASX (5′-GAATTCGATTCGGGATCCX-3′ where X is G, C, A, or T) directed amplification of a 98 bp allele-specific amplicon from the plasmid DNA template. The 3′-terminus of primer PMASX was designed to anneal to the single polymorphic base within the plasmid sequence. Product melting temperatures were predicted as 85.0 °C using PrimerCalc software, which employs nearest neighbour thermodynamic algorithms for T_m calculations [28].

Reactions were performed in a total volume of 20 μl containing 2 μl LightCycler DNA FastStart SYBR Green I mixture (*Taq* DNA polymerase, reaction buffer, dNTPs, and SYBR Green I dye; Roche Applied Science, Lewes, UK), a final concentration of 3 mM $MgCl_2$, and PCR-grade water. Forward and reverse primers were present at 0.5 M each with 100 pg plasmid template. Each forward primer (PMASX) with PMOP98 was used to direct amplification from four plasmid templates (containing G, C, A, or T variant nucleotide) testing all 16 base combinations. Reactions were performed in triplicate with a single control sample void of template.

Continuous PCR monitoring and melt curve analysis. PCR cycling and melt analysis protocols were performed on the LightCycler PCR instrument (Roche Applied Science). Amplifications were completed in approximately 40 min, including a 10-min *Taq* DNA polymerase activation step at 95 °C, 50 PCR cycles of denaturation at 95 °C for 1 s, primer annealing at 65 °C for 5 s, and primer extension for 5 s at 72 °C. All temperature transition rates were set to 20 °C/s and each PCR cycle was monitored by continual fluorescence acquisition. Fluorescence gains (used to set sensitivity of detection) for the LightCycler were F1-1 throughout. These continuous monitoring data were exported into Microsoft Excel 2000 for further analysis.

Reaction specificity was confirmed by executing a melt curve analysis immediately after amplification in the same closed capillary tube. An initial denaturation at 95 °C was followed by a cooling step to 70 °C (5 °C above primer annealing) and a slow denaturation phase to 95 °C at a rate of 0.1 °C/s, with continual fluorescence acquisition. Product denaturation was observed as a rapid loss of fluorescence near the calculated T_m, and converted to a melt peak by plotting the first negative derivative of fluorescence loss, $F1$, with respect to temperature, T ($-\mathrm{d}(F1)/\mathrm{d}T$ vs T). A single peak at the predicted T_m confirms the optimal nature of the reaction conditions and the absence of contamination in control samples. A broad peak at a significantly lower T_m than PCR product identifies primer dimer amplification [29].

Collection of 'per cycle' rate data. The reaction rate v was determined from the maximal increase in fluorescence signal $F1$ with respect to time, t, at 72 °C at each cycle. The rates were estimated from the linear portion of the progress curve, as shown in Fig. 1C, according to the linear equation,

$$F1 = vt + c, \tag{1}$$

where c is the $F1$ reading at time zero. The slope v of fluorescence signal increase was assumed proportional to the rate of nucleotide incorporation by *Taq* DNA polymerase since this will be concomitant with the intercalation of SYBR Green I molecules into the newly synthesised double-stranded DNA.

Results and discussion

Kinetic data

Fig. 1A shows a typical plot of fluorescence $\Delta F1$ during PCR. The observed transitions in fluorescence closely follow the temperature profile because of the strong dependency of fluorescence on temperature [4]. As the sample is heated, fluorescence is high until denaturation occurs (apparent as a sharp drop in fluorescence). As the sample cools from denaturation to annealing temperature, fluorescence increases rapidly, reflecting product-to-product annealing (Fig. 1B). Fluorescence also increases during extension whilst the

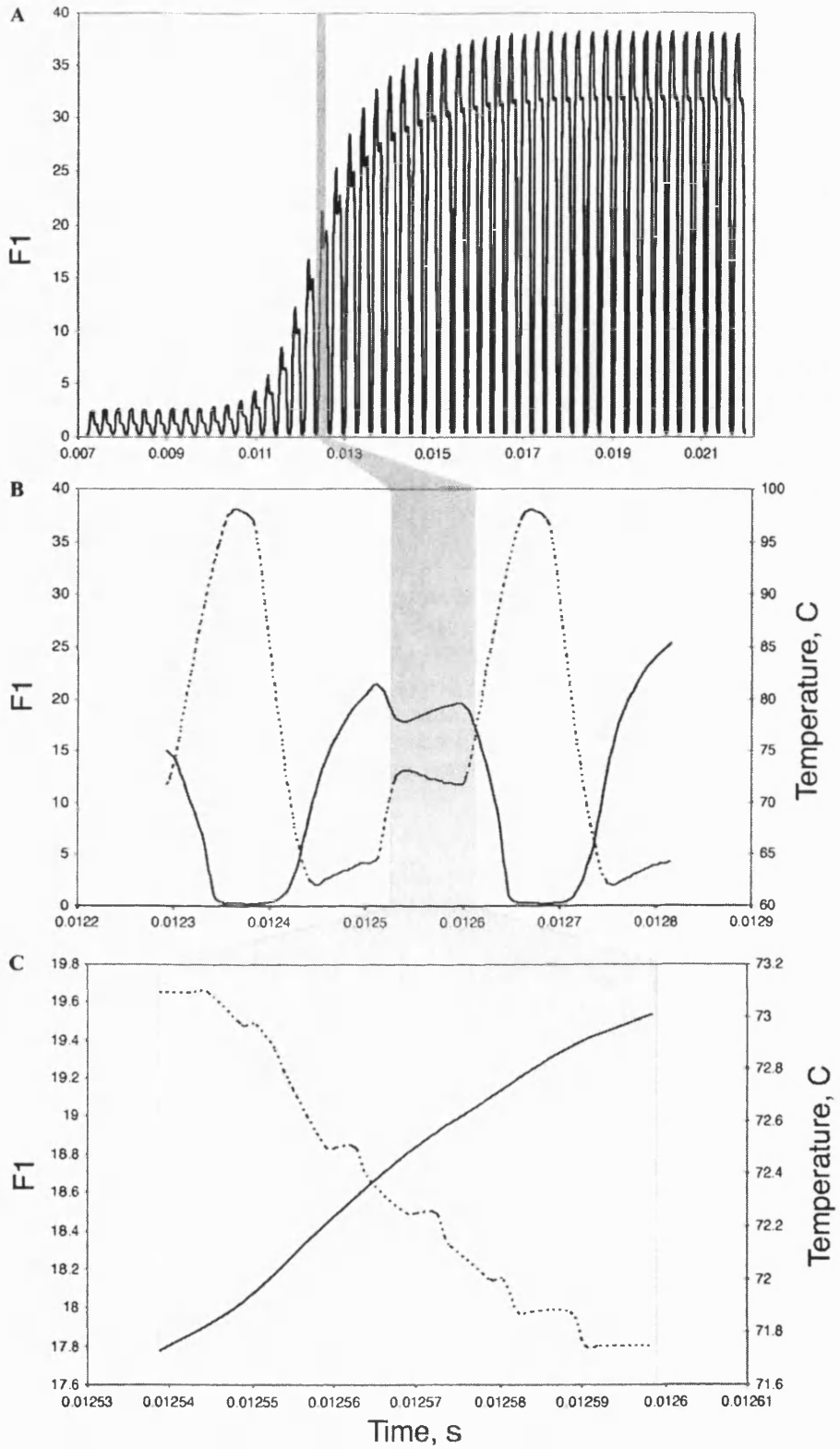


Fig. 1. Grey section highlights area of interest from a matched allele-specific PCR test from template containing variant nucleotide T. (A) Continuous monitoring data over 50 PCR cycles. (B) Temperature dependence of product accumulation over a single cycle. (C) Product accumulation during the extension phase (72 °C) only.

temperature is held within the optimal limits for *Taq* DNA polymerase (Fig. 1C). This increase is directly attributable to polymerisation and the accumulation of double-stranded DNA. During early and late cycles fluorescence changes over this period were negligible. This is in contrast to cycles where exponential and linear amplification of target would be expected and where significant increases in fluorescence were observed from the real-time data.

Mathematical model

The rate of reaction at each cycle was calculated from the slope of the linear $F1$ portion in Fig. 1C. Data points in Figs. 2A–D show the cycle dependence of rate v using template with base variant 'T'. The lag phase during early cycles was attributable to fluorescence signals that were below the detection limits of the optics in the LightCycler. The length of the lag phase varied according to base complementation and was generally less for a matched template compared to a mismatch. This lag phase was followed by a rapid increase in reaction velocity, which reached a maximum, and then decreased as cycle number was increased further.

The dependency of reaction rate on substrate concentration shows a strong departure from the hyperbolic relationship predicted by the Michaelis–Menten equation, and appears to be characteristic of high substrate inhibition in which the formation of abortive complexes inhibits the enzyme. The reaction rates generated from real-time data closely followed the general curve predicted for high substrate inhibition,

$$v = \frac{V_{\max}}{[1 + (K_M/S) + (S/K_i)]}, \tag{2}$$

where $V_{\max}$ is the maximum velocity, K_M is the Michaelis–Menten constant, and K_i the inhibition constant. At low substrate concentration S, (S/K_i) is insignificant compared to $1 + (K_M/S)$ and Eq. (2) becomes identical to the classic Michaelis–Menten equation. It follows that as S increases, the value of (S/K_i) increases to become the significant term in the denominator and, as observed in our experiments, the value of v decreases [27].

Since we fit this equation to data generated from PCR, a non-classical enzyme system, S must be represented appropriately. In PCR the product formed in cycle $n-1$ is the substrate that is available for cycle n. Under non-limiting conditions this should result in an exponential increase in substrate concentration over successive cycles described Eq. (3),

$$S_n = 2^n S_0, \tag{3}$$

where S_0 is the initial template (substrate) concentration and S_n is the substrate concentration after n successive cycles. It is well understood, however, that in PCR this

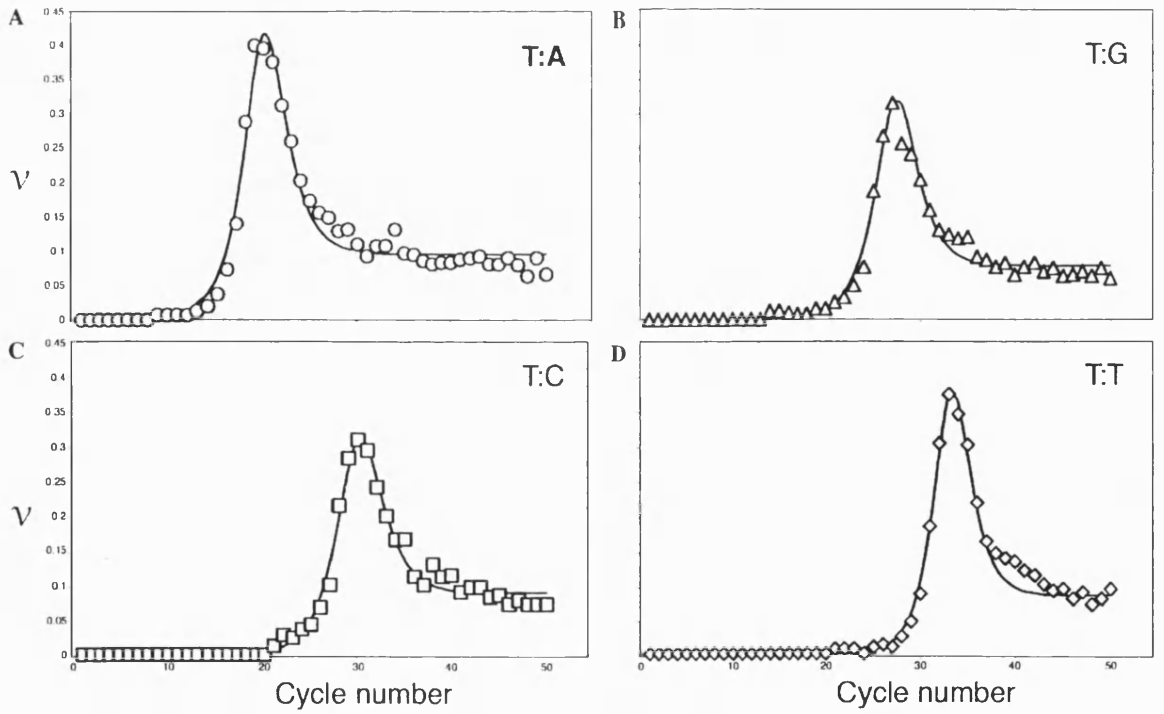


Fig. 2. Graphs of rate v vs. cycle number for base combinations (template: primer) (A) T:A, (B) T:G, (C) T:C, and (D) T:T. The solid line represents the curve predicted by the model (Eq. (5)) and points are data collected from real PCR tests using Eq. (1). Correlation coefficient r^2 is >0.99 in each case.

model holds only for early cycles [20]. We applied a simplified sigmoidal function to describe the accumulation of substrate during PCR,

$$S = \frac{1}{(1 + a\exp^{(-fn)})} \tag{4}$$

where a and f are descriptive parameters of the sigmoid function at cycle n [30]. This function arises in many dynamical systems, describing simple exponential growth dynamics with a linear limiting control. Such systems are often called logistic growth. A similar sigmoidal function was recently reported as providing a good fit for the whole kinetic process of real-time PCR [31]. The inclusion of Eq. (4) as a substrate function approximated well to real-time data in our experiments (Figs. 3A–D), particularly in correctly primed PCR tests.

Our model was further refined by the inclusion of a compensation factor C. Eq. (2) demands that at zero substrate concentration the rate of the reaction is zero. Since a PCR contains an initial substrate (template) concentration at cycle number $n = 1$, this compensation factor allows the criteria of the equation to be satisfied, such that an artificial cycle 0 is included. The overall equation becomes,

$$v = \frac{V_{\max}}{[1 + (K_M/(S + C)) + ((S + C)/K_i)]}, \tag{5}$$

where S is represented by the sigmoidal function of Eq. (4).

Determination of kinetic constants in allele-specific PCR

The synthesis of oligonucleotides that differ only at the 3′-terminal base, coupled with templates with a variant nucleotide at a defined position, provided a system to evaluate the effects of mismatches on PCR. Non-linear regression (NL-REG Version 5.4; Phillip H. Sherrad) was used to estimate informative kinetic parameters for each base configuration using Eq. (5). The ability of *Taq* DNA polymerase to bind and extend terminal mismatches was dependent upon the identity of the mispair (Table 1).

We assumed that V_{max} represented the maximal rate of incorporation of SYBR into an extending DNA complex under non-limiting conditions. This provides an arbitrary measure of maximal polymerisation or catalytic activity. Although V_{max} is not a fundamental property of an enzyme as it is dependent on enzyme concentration [25], we were able to compare reactions using this parameter since the same total enzyme concentration was used in each PCR test.

For amplifications from template with a variant nucleotide G, V_{max} was highest in the presence of a matched primer 3′ terminus. For all other templates at least one of the mismatched primers gave a higher V_{max} value than the match. In PCR tests involving the transversion

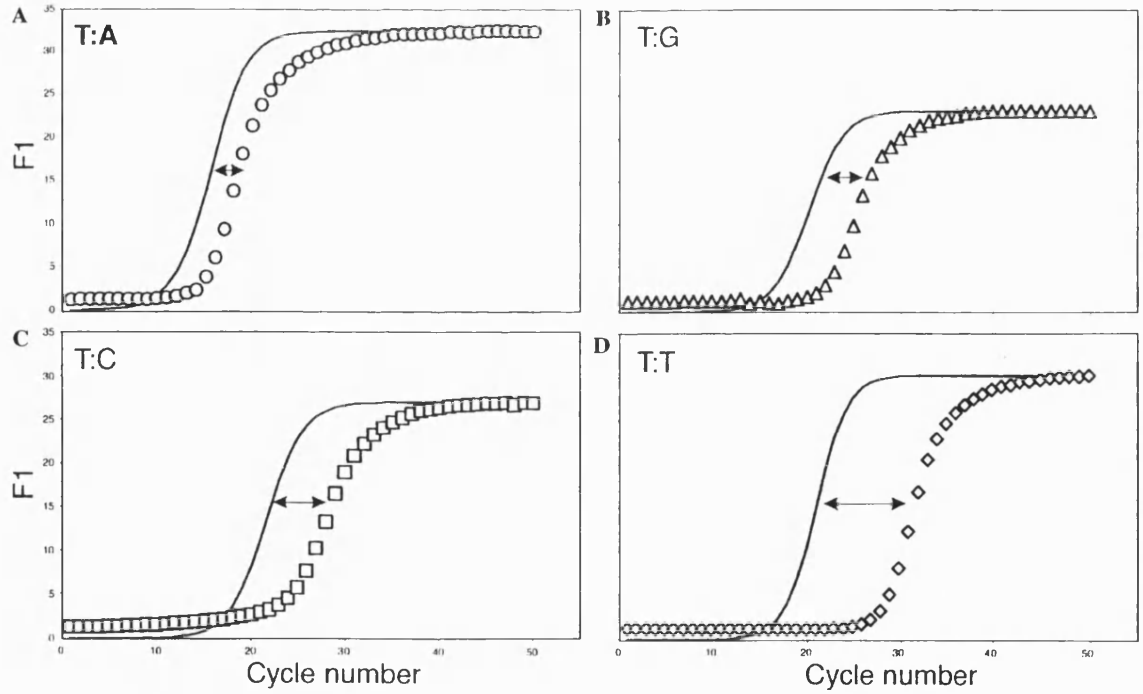


Fig. 3. Correlation of real-time substrate accumulation (data points) with that predicted by the sigmoid function multiplied by final fluorescence reading (solid line). Graphs display base combinations (template: primer) (A) T:A, (B) T:G, (C) T:C, and (D) T:T. Arrows indicate the larger deviations from predicted values in the presence of a mismatch.

Table 1
Predicted kinetic parameters (shown to 3 s.f.) for each base configuration

Template base	Primer 3′ terminus	V_{max}	K_M	K_i	V_{max}/K_M	Ratio of V_{max}/K_M[a]
G	**C**	**530**	**1300**	**0.002**	**0.408**	**1.000**
	G	448	43100	0.083	0.010	0.025
	A	88.1	479000	29.2	0.000	0.004
	T	302	23600	0.111	0.013	0.031
C	C	290	14700	0.046	0.020	0.106
	G	**335**	**1800**	**0.006**	**0.186**	**1.000**
	A	378	26100	0.093	0.015	0.078
	T	271	25400	0.128	0.011	0.058
A	C	906	49900	0.037	0.018	0.089
	G	402	32500	0.082	0.012	0.061
	A	45	1720000	369	0.000	0.000
	T	**377**	**1840**	**0.007**	**0.204**	**1.000**
T	C	475	17700	0.031	0.027	0.120
	G	669	16700	0.015	0.040	0.180
	A	**245**	**1090**	**0.013**	**0.224**	**1.000**
	T	160	116000	2.63	0.001	0.006

[a] Ratio of specificity constants compared to a matched combination. Matched combinations are highlighted in bold type. Constants are displayed in arbitrary units.

mispairs (template:primer) G:A, A:A, and T:T, V_{max} was lower than that found for a match. The transition mispairs C:A, A:C, and T:G and transversion mispairs A:G and T:C exhibited equal or higher V_{max} than a matched combination.

The latter observation is not surprising when one considers the consequence of mismatched primer extension in PCR. Certain nucleotide configurations vary in their affinity to bind to the enzyme and extend. However, once extension from a mismatched primer occurs, the resultant product and the complement synthesised in subsequent cycles are fully matched with both primers [6]. The nature of PCR means that a single copy of matched template can be amplified to plateau levels in just a few cycles. This causes failure in many end-point genotyping tests by impairing discrimination. Therefore, V_{max} alone is insufficient to ascertain the true effects of a mispair in PCR.

The more dominant parameter for characterising enzymatic discrimination is the Michaelis constant, K_M, defined as the substrate concentration at which the reaction rate is half its maximal velocity [13]. The K_M is an intrinsic property of an enzyme reflecting the binding constant for forming the enzyme substrate complex, as well as the catalytic constant. In PCR, this embodies template affinity and the thermodynamic environment of the reaction [20]. The transversion mispairs G:A, A:A, and T:T exhibited a K_M value between 100- and 1000-fold higher than a corresponding match. In all other cases, K_M for a mismatch was 8- to 33-fold higher than the equivalent match. These data suggest that a matched conformation binds with higher affinity than a mismatch to the active site of the enzyme.

This is discordant with the view of Huang et al. [5] who found that *Taq* polymerase binds with equal affinities to all base configurations. This is likely due to the reaction conditions employed where a single species of dNTP was used as the variable substrate for each template. Single-turnover kinetics is unable to reflect a PCR test in which all dNTPs are present at equimolar concentration and in large excess. In the absence of competition, similar binding affinities might be expected. Some groups were similarly conflicting [10,14], whilst others were in agreement [7,9,13,15]. The diversity of reaction conditions employed, enzyme source, and DNA sequences may have caused this disparity. Importantly, this report represents the kinetic characterisation of *Taq* in a true rapid cycle PCR environment overcoming the limitations of previous findings.

The kinetic parameters displayed in Table 1 were used to define a value for the fundamental 'specificity' constant, k_{cat}/K_M, where k_{cat} is the catalytic constant or turnover number. This specificity constant determines the ratio of reaction rates for an enzyme acting on two competing substrates, when they are mixed together at equimolar concentrations. Given the fact that it is the ratio of specificity constants that determines the ratio of rates of the competing reactions a and b, we evaluated V_{max}/K_M since $[E_t]$ was identical throughout our tests [25]

$$\frac{(V^A_{max}/K^A_M)a}{(V^B_{max}/K^B_M)b} = \frac{((V^A_{max}/E^A_T)/K^A_M)a}{((V^B_{max}/E^B_T)/K^B_M)b} = \frac{(k^A_{cat}/K^A_M)a}{(k^B_{cat}/K^B_M)b} \quad (\text{if } E^a_T = E^b_T). \tag{6}$$

The comparison of these values between matched- and mismatch-primed PCR expresses the enzyme's ability to discriminate in favour of a particular base configuration in the presence of others [25], and consequently, the relative extension efficiencies [5]. In agreement with work by Kwok et al. [6] we found that a match was the more specific substrate for *Taq* polymerase. The relative specificity of mismatches varied according to each arrangement, but in general, purine:purine mispairs were harder to extend than other combinations. Our system unequivocally demonstrates that for *Taq* DNA polymerase during rapid cycle PCR, the matched base configuration would be the preferred substrate in the presence of all other mismatches.

Finally, this model estimated values for the inhibition constant K_i, which provides a quantitative measure of inhibitor potency [27]. Values for K_i are usually derived from an inhibition study of the effects of a substance on kinetics or binding of a substrate or effectors. Such studies are usually performed in the presence of the true substrate (for kinetics) or ligand (for binding) AND the inhibitor. Therefore, the K_i value in this model represents a substrate inhibition constant since the substrate only is present. This could be interpreted as the ability of the mismatched primer to form a dead-end complex by binding to the active site of the enzyme without efficient extension. In addition, it may represent the ability of the primers to bind dystopically to a particular template, such that replication rate is depressed. In both cases, one might expect K_M for a mismatched primer to be higher than that of the true primer. This is supported by data in which the three transversion mispairs, G:A, A:A, and T:T, exhibited a significantly higher K_i value than all other mispairs. Consistently, K_i was lower for a match. By considering the other parameter values associated with each configuration, we propose that these particular mispairs bind to the enzyme with low affinity and extend with low efficiency, forming a stable dead-end complex that inhibits progression of the reaction.

High substrate inhibition

Factors that have been attributed to attenuation of PCR include depletion of substrate (dNTPs or primers), thermal inactivation or limiting concentration of DNA polymerase, inhibition of enzyme activity by increasing pyrophosphate production, reannealing of amplicon at concentrations above 10^{-8} M, reduction in the denaturation efficiency per cycle, destruction of product due to enzyme 5′–3′ exonuclease activity [32], product-to-product reannealing [4] or the chelation of critical metal ions by the substrate depriving the enzyme of a cofactor [27]. For each theory, the characteristics of high substrate inhibition would be apparent.

Other investigations have shown that by including increasing amounts of "random" DNA into the PCR test from the beginning amplification could be inhibited [32,33]. It was suggested that the accumulation of product during later cycles is more likely to inhibit the enzyme before the more trivial factors such as exhausting primers, metal ions or dNTPs, which are added in huge molar excess. In conjunction with the excellent fit of our model to real data (Figs. 2A–D), we provide compounding evidence that high substrate inhibition does indeed play a predominant role in curtailing amplification during the final cycles of a PCR.

However, one might have expected the K_i for a match to be greater than or equal to a mismatch if we were proposing high substrate inhibition as the sole cause of the PCR plateau phase. Our data show that in the case of a mismatch, other factors come into play, which further add to the magnitude of the K_i value. This is supported by Figs. 3A–D in which the model's prediction of substrate accumulation was more accurate for a match- than mismatch-primed PCR test. So although high substrate inhibition is the likely cause of the plateau phase in conventional PCR, we demonstrate that the actual presence of a mismatch has an effect on the intrinsic ability of the enzyme to function effectively, displaying a direct inhibitory effect on the PCR.

In summary, we report for the first time kinetic parameters that have been estimated in a real-time rapid PCR system. These data represent the enzyme's activity under optimal reaction conditions. Due to the complex interactions of components within a PCR [34], and the thermodynamics of DNA hybridisation [13], variations in component concentrations or annealing temperature during the test will alter the resulting kinetic constants. It is, therefore, important to only compare kinetic parameters obtained under identical reaction conditions.

The general trends shown here are directly applicable to the design of SNP genotyping experiments, providing a measure of the likely success and optimisation necessary. We expect this novel approach to PCR characterisation to be utilised in other applications, especially rapid qPCR methods in which accurate measurement of quantitative parameters is essential.

Acknowledgment

This work was supported by research grants from the University of Bath and Molecular Sensing plc.

References

[1] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, Cold Spring Harbour Symp. Quant. Biol. 51 (1986) 263–273.

[2] R.K. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, N. Arnheim, Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230 (1985) 1350–1354.

[3] S.A. Bustin, Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, J. Mol. Endocrinol. 25 (2000) 169–193.
[4] C.T. Wittwer, M.G. Herrmann, A.A. Moss, R.P. Rasmussen, Continuous fluorescence monitoring of rapid cycle DNA amplification, BioTechniques 22 (1997) 130–138.
[5] M.-M. Huang, N. Arnheim, M.F. Goodman, Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR, Nucleic Acids. Res. 20 (1992) 4567–4573.
[6] S. Kwok, D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, J.J. Sninsky, Effects of primer template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies, Nucleic Acids Res. 18 (1990) 999–1005.
[7] J.P. Day, D. Bergstrom, R.P. Hammer, F. Barany, Nucleotide analogs facilitate base conversion with 3′ mismatch primers, Nucleic Acids Res. 27 (1999) 1810–1818.
[8] C. Christopherson, J. Sninsky, S. Kwok, The effects of internal primer template mismatches on RT-PCR: HIV-1 model studies, Nucleic Acids Res. 25 (1997) 654–658.
[9] L.V. Mendelman, J. Petruska, M.F. Goodman, Base mispair extension kinetics: comparison of DNA polymerase α and reverse transcriptase, J. Biol. Chem. 265 (1990) 2338–2346.
[10] L.B. D'yachenko, A.A. Chenchik, G.L. Khaspekov, A.O. Tatarenko, R.S. Bibilashvili, Efficiency of DNA synthesis initiated by complementary and mismatched primers, Mol. Biol. 28 (1994) 654–660.
[11] S. Ayyadevara, J.J. Thaden, R.J. Shmookler Reis, Discrimination of primer 3′-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction, Anal. Biochem. 284 (2000) 11–18.
[12] C.R. Newton, A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, A.F. Markham, Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS), Nucleic Acids. Res. 17 (1989) 2503–2515.
[13] J. Petruska, M.F. Goodman, M.S. Boosalis, L.C. Sowers, C. Cheong, I. Tinoco, Comparison between DNA melting thermodynamics and DNA polymerase fidelity, Proc. Natl. Acad. Sci. USA 85 (1988) 6252–6256.
[14] R.D. Kuchta, V. Mizrahi, P.A. Benkovic, K.A. Johnson, S.J. Benkovic, Kinetic mechanism of DNA polymerase I (Klenow), Biochemistry 26 (1987) 8410–8417.
[15] I. Wong, S.S. Patel, K.A. Johnson, An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics, Biochemistry 30 (1991) 526–537.
[16] C.T. Wittwer, B.C. Marshall, G.H. Reed, J.L. Cherry, Rapid cycle allele-specific amplification: studies with the cystic fibrosis ΔF508 locus, Clin. Chem. 39 (1993) 804–809.
[17] J. Peccoud, C. Jacob, Statistical estimations of PCR amplification rates, in: Gene Quantification, Birkhauser, Boston, 1998.
[18] G. Stolovitzky, G. Cecchi, Efficiency of DNA replication in the polymerase chain reaction, Proc. Natl. Acad. Sci. USA 93 (1996) 12947–12952.
[19] M.V. Velikanov, R. Kapral, Polymerase chain reaction: a Markov process approach, J. Theor. Biol. 201 (1999) 239–249.
[20] S. Schnell, C. Mendoza, Theoretical description of the polymerase chain reaction, J. Theor. Biol. 188 (1997) 313–318.
[21] S. Schnell, C. Mendoza, Closed form solution for time-dependent enzyme kinetics, J. Theor. Biol. 187 (1997) 207–212.
[22] L. Raeymaekers, Quantitative PCR: theoretical considerations with practical implications, Anal. Biochem. 214 (1993) 582–585.
[23] G. Weiss, A. Haeseler, A coalescent approach to the polymerase chain reaction, Nucleic Acids Res. 25 (1997) 3082–3087.
[24] S. Schnell, C. Mendoza, Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction QC-PCR, J. Theor. Biol. 184 (1997) 433–440.
[25] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, Butterworth, London, 1979.
[26] A.J. Brookes, The essence of SNPs, Gene 234 (1999) 177–186.
[27] C.W. Wharton, R. Eisenthal, Molecular Enzymology, Blackie, London, 1981.
[28] K.J. Breslauer, R. Frank, H. Blocker, L.A. Marky, Predicting DNA duplex stability from the base sequence, Proc. Natl. Acad. Sci. USA 83 (1986) 3746–3750.
[29] K.M. Ririe, R.P. Rasmussen, C.T. Wittwer, Product differentiation by analysis of DNA melting curves during the polymerase chain reaction, Anal. Biochem. 245 (1997) 154–160.
[30] D.H.V. Seggern, CRC Standard Curves and Surfaces, CRC Press, Boca Raton, 1993.
[31] W. Liu, D.A. Saint, Validation of a quantitative method for real time PCR kinetics, Biochem. Biophys. Res. Commun. 294 (2002) 347–353.
[32] P. Kainz, The PCR plateau phase—towards an understanding of its limitations, Biochim. Biophys. Acta 1494 (2000) 23–27.
[33] D.S. Dimitrov, M.A. Apostolova, The limit of PCR amplification, J. Theor. Biol. 178 (1996) 425–426.
[34] U. Linz, U. Delling, H. Rübsmaen-Waigmann, Systematic studies on parameters influencing the performance of the polymerase chain reaction, J. Clin. Chem. Clin. Biochem. (1990) 5–13.